DSI DECISION SCIENCES INSTITUTE

SCHOLARS • EDUCATORS • PROBLEM SOLVERS

2023 DSI ANNUAL CONFERENCE

PROCEEDINGS

Responsive, Relevant, Responsible: Decision Making in the Age of Transformation

November 18 – 20, 2023

Marriott Marquis Atlanta, GA (USA)

Download at:
https://decisionsciences.org/wp-content/uploads/FINAL_PROCEEDINGS_2023_v4.pdf

ISBN: 979-8-9897032-0-3
ISSN: 2471-884x

TABLE OF CONTENTS

277 **Case-based Reasoning for Meta-learning of Visual Inspection Algorithm Selection**
Shu Wang (Georgia Institute of Technology), Mulang Song (Mr.), Yiyun Fei (Georgia Institute of Technology), Roger Jiao (Georgia Institute of Technology)

287 **Climate Change Risk Assessment for a University Campus**
John Maleyeff (Boston University), David Weidman (Boston University), Jiaoxue Liu (Boston University), Maria Kristina (Boston University), Yujue Tan (Boston University)

301 **Closet-Building vs. Minimalism: Selling Fewer, Better Products to Fashion-Sensitive Customers**
Monire Jalili (Bentley University), Michael Pangburn (University of Oregon), Alireza Yazdani (Cal Poly Pomona)

302 **Collaborative New Product Development: Co-creating with a Shared Supplier in the Presence of a Competitor**
Abhishek Roy (Temple University)

3 **Cookie & Card Game: Understanding Machine Learning Algorithms Using Humanizing Pedagogy**
Sudipendra Nath Roy (Herberger Business School, St. Cloud State University)

3 **COVID-19 Uncertainty on Corporate Risk-taking: International Evidence**
Md Ismail Haidar (University of Texas Rio Grande Valley)

305 **Diabetes Mellitus Trend among Adults in Texas Counties from 1999-2009: A Spatial Analysis of Events**
Theresa Abah (California State University)

321 **Discretion in Automated Supermarket Replenishment: Censorship Bias and Self-inflicted Stockouts**
Bengu Nur Ozdemir (IE Business School), Antti Tenhiälä (IE Business School)

322 **Eco-efficiency Assessment of the OECD Countries by DEA Models**
Umit Saglam (East Tennessee State University)

336 **EDI Utilization in Pharmacy Revenue Cycle and the Prioritization of Exploitation over Exploration**
Derek Dubois (The University of Rhode Island), Mehmet Gokhan Yalcin (The University of Rhode Island College of Business)

337 **Effects of Gamification Training on Employees' Information Security Compliance**
Venkata Kumar Suram (University of the Cumberlands), Mary L Lind (LSU Shreveport)

363 **Emotion Detection in Texts: Efficiency Comparison of Different Classifiers**
Debangi Choudhury (Ravenshaw University, Cuttack, Orissa, India), Meera Behera (Georgian Court University)

378 **Environmental Policy and Corporate Default Risk: International Evidence**
Md Ismail Haidar (University of Texas Rio Grande Valley)

Excel-Based Continuous Review Inventory Game
Jay Brown (Loyola University Maryland), Maxim Bushuev (Morgan State University), Tatiana Rudchenko (Georgia Institute of Technology)

1063 **The Critical Success Factors and Barriers of Lean in Small and Medium-Sized Enterprises: A Systematic Literature Review**
Kubra Gurtas (Middle East Technical University), Arsev Umur Aydinoglu (Middle East Technical University), Fatma Pakdil (Eastern Connecticut State U)

1095 **The Effects of Personalized Push Notifications on Mobile Shoppers' Attitude Towards Push Notification Engagement and Acceptance.**
Mbuso Levon Gama (National Cheng Kung University), Jengchung Victor Chen (National Cheng Kung University)

1118 **The Impact of Process Degradation on Improvement Activity: An Exploratory Study**
Brad C Meyer (Drake University), Dan Bumblauskas (Missouri Western State University), Richard Keegan (Trinity College Dublin Business School), Dali Zhang (Shanghai Jiao Tong University, Antai College of Economics and Management)

1136 **The Impact of Supply Chain Strategy and Entrepreneurial Orientation on Supply Chain Resilience**
Prashanth Anekal (Saginaw Valley State University), Surender Reddy (Saginaw Valley State University)

1 2 **The Importance of Project Management Ethics and Its Effects on Business Success**
Maling Ebrahimpour (The University of Rhode Island), Chrsitopher Orr (The University of Rhode Island), Rachel Vennel (The University of Rhode Island), Paulina Loredo-Vilchis (The University of Rhode Island)

1 3 **The influence of power dynamics on e-mail communication**
Joy Younsoo Kim (Columbia University in the City of New York)

1163 **The Intellectual Structure of Social Engineering Research: A Co-occurrence Approach** *Anil Singh (University of Texas at Arlington), George Mangalaraj (Western Illinois University), Aakash Taneja (Stockton University)*

1176 **The Power of Success: How High-Performing Organizations Navigate Major Transformations Without Sacrificing Credibility**
Majid Majzoubi (York University)

1177 **The role of leadership styles, organizational culture, and knowledge management in higher education institutions (HEI)**
Armel Djangone (Dakota State University), Omar El-Gayar (Dakota State University)

1199 **The Standardized Development Strategy of Shanghai Community Sports Volunteer Service**
Xinyu Wu (Tongji University), Tianfeng Lu (Tongji University)

1211 **Toward a Conceptual Framework of International Tourism after the Pandemic: Consumption Propensity Based on Expectations of Refreshment and Development**
Seung Hoon Jang (Commonwealth University of Pennsylvania), Christian Grandzol (Commonwealth University of Pennsylvania)

1218 **Toward a Normative Framework for Systemic Risk Modeling**
Faraz Dadgostari (Montana State University)

1 2 **When Is Genetic Engineering Going Too Far?**
Aurore Jose Kamssu (Tennessee State University), Eric Philip (Tennessee State University), Mustafa Alsadeq (Tennessee State University), Ali Nasib Nasib (Tennessee State University)

13 2 **Youth Narratives In Relation to Trauma**
Audra E Dankwardt (UMass Amherst), Jesslynn Rocha Neves (University of Massachusetts)

1401 **Zooming in on Inflation and Associations with Market Returns**
Raymond Yu Shao (William Mason High School)

DECISION SCIENCES INSTITUTE

A Blockchain Ecosystem for Safer Circular Lithium-ion Battery Supply Chains: A Performance-based Theoretical Perspective

Zhuowen Chen
Worcester Polytechnic Institute
Zchen14@wpi.edu

Abdullah Yildizbasi
Worcester Polytechnic Institute
ayildizbasi@wpi.edu

Joseph Sarkis
Worcester Polytechnic Institute
jsarkis@wpi.edu

ABSTRACT

The circular lithium-ion battery (LIB) supply chain presents challenges in safety, transparency, and efficiency. Blockchain technology offers the potential to enhance safety management by tracking battery movement. However, a research gap exists in applying blockchain for safety in this circular supply chain. This necessitates exploring blockchain's role in safety, guided by strong theoretical frameworks. To address this gap, we propose a theory-supported blockchain application for the circular LIB supply chain, aiming to improve safety. Our approach involves a comprehensive blockchain ecosystem with an architecture tailored for safety. It emphasizes stakeholder engagement and a safety measurement matrix. Theoretical-based propositions provide insights for research and practical blockchain use. This study highlights the need for theoretical exploration, refined blockchain architecture, and safety prioritization for a secure circular economy. Empirical investigation is essential, outlining areas like safety performance assessment. Practical implications span businesses, policies, and circular economy advancement. While this research lays the groundwork for applying blockchain to circular supply chain safety, limitations exist due to spatial constraints. Specific inquiries, like blockchain implementation algorithms, were not addressed. The blockchain architecture needs further development. In conclusion, our research contributes by proposing blockchain for circular supply chain safety management. The model integrates practical strategies and theoretical insights, paving the way for blockchain in LIB safety and beyond.

KEYWORDS: Blockchain, Lithium-ion battery, Stakeholder theory, Circular economy, Performance measurement

DECISION SCIENCES INSTITUTE

Closing the Gap: Enhancing Geographic Equity under the Continuous Distribution Organ Allocation Policy

Shubham Akshat
Carnegie Mellon University, Pittsburgh
Email: sakshat@andrew.cmu.edu

S. Raghavan
University of Maryland, College Park
Email: raghavan@umd.edu

Nicholas L. Wood
Hennepin Healthcare Research Institute, Minneapolis
Email: nick.wood@cdrg.org

ABSTRACT

The deceased-donor organ allocation system is going through a major overhaul in the United States. Going forward, all the organ allocation systems (including the liver) will be based on a Continuous Distribution framework. We develop a novel analytical method to model the supply (deceased donors)-to-demand (waiting list patients) ratio at a transplant program. We integrate it into a set-partitioning optimization model to design an equitable allocation policy under the new policy framework. Using a simulation model, we illustrate the benefits of our model in achieving geographic equity over the current and other "one-size-fits-all" type policies.

KEYWORDS: Liver allocation, Healthcare policy, Geographic disparity, Applied probability, and Operations research

DECISION SCIENCES INSTITUTE
Green Crowdsourcing Mobile Application for Groceries in South Africa

Marcia **Mkansi**
University of South Africa
Email: mkansm@unisa.ac.za

John Michael Maxel **Okoche**
University of South Africa
Email: okochjm@unisa.ac.za

Mangaliso **Sipamla**
University of South Africa
Email: 57661847@unisa.ac.za

ABSTRACT

This study presents a green crowdsourcing mobile application (App) innovation, named Dilivari, which responds to a multi-embedded groceries supply chain coordination problem linked to people, profitability, and environmental management in South Africa. The application was developed using a qualitative computer programming approach, informed by a design science methodology. The computer experimental research strategy reveals a functional mobile application that responds to the needs of the multi-embedded supply chain coordination problem related to the groceries. Theoretically, the innovation extends existing crowdsourcing frameworks beyond what might normally be expected of coordination theories and includes profitability, reduction of congestion and environmental sustainability.

KEYWORDS: Green Crowdsourcing App, Grocery Supply Chain, Supply Chain Coordination, Design Science

INTRODUCTION

The importance of the development of an appropriate crowdsourcing application (App) for groceries can never be understated as it contributes towards balancing people, profitability, environment, and economic development. It is the 13th highest sustainable development goal (SDG 13) of the United Nations (UN) and is widely recognised by policymakers, professional funding bodies and scholars as one of the critical, interminable, and intractable global challenges that require innovation and cohesive efforts from all community stakeholders (Mkansi, 2013). The Dilivari App provides a collaborative process for meeting requirements for the distribution of groceries in South Africa. The innovation solves problems of poverty alleviation together with supporting environmental sustainability. The App solves contemporary problems in developing countries innovatively and creatively by collecting information using a crowdsourcing approach. As an improvement to the existing crowdsourcing Apps the critical aspects of congestion and environmental management have been incorporated (Stol & Fitzgerald, 2014). This will be useful in Africa, where some of the greatest challenges for growth and development are unemployment, environmental management, inadequate capacity to coordinate groceries, warehouses, households, industry, academia, and corporations' contribution to a shared agenda towards the transformation. The role of every segment of society, including academics and industry is imminent because governments' budgets are shrinking and non-government bodies are grappling with limited funds post Covid-19, which wreaked havoc in most African economies. Recognising the transformational power of innovation and social-embedded solutions in creating social value for people of all age groups is a global opportunity recognised by both the African Union Agenda 2063 and UN priority (African Union, n.d).

LITERATURE REVIEW

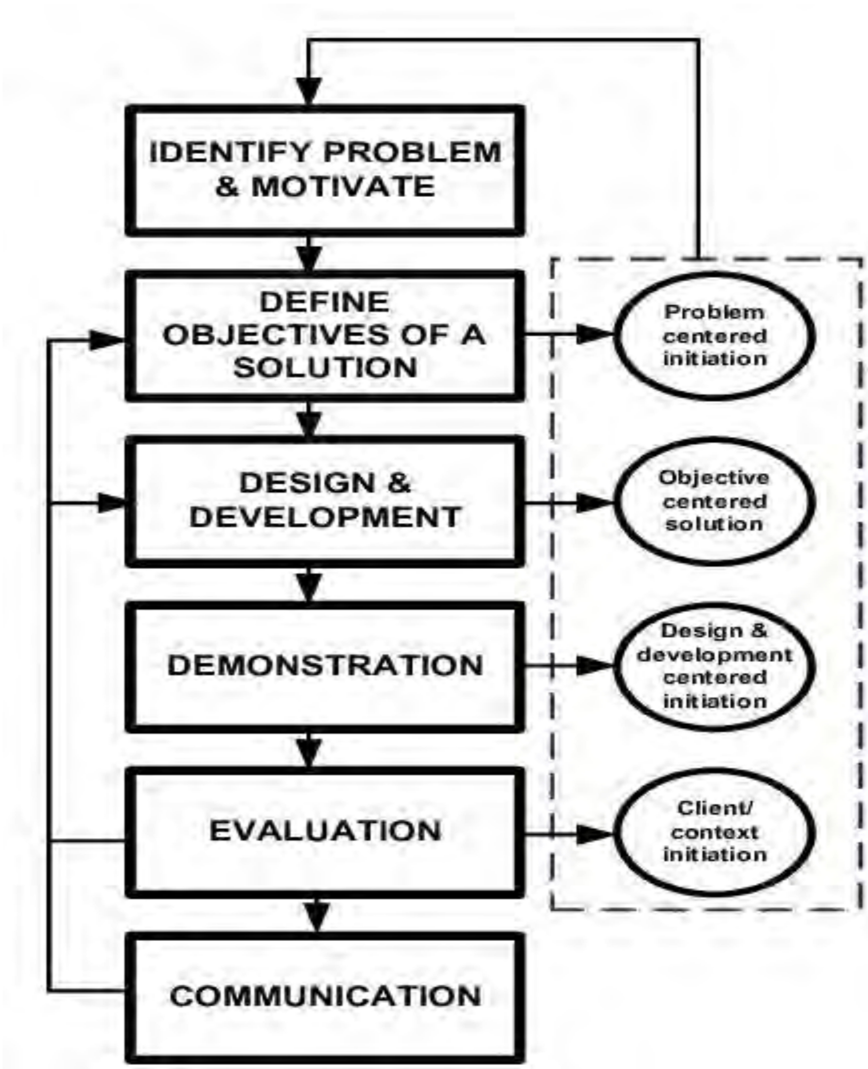
The green crowdsourcing App provides a collaborative process (Beck et al., 2022), problem-solving (Bassi et al., 2020), creativity, innovation, knowledge capture (Franzoni et al., 2022), collection of information (Edgar et al., 2016) or data (Sheehan, 2018), theory development (Mason & Suri, 2012), support in the identification of new research problems (Beck et al., 2022), interactions among group members leading to knowledge development (Law et al., 2017) and workflow tasks, assets, processes and outputs (Hedges & Dunn, 2017). Despite the appreciation of apps, balancing the triple effect of people, profitability and the environment is critical towards the advancement of theory and practice in grocery transportation. Mkansi (2013) addressed the theoretical problem of balancing the triple effect of people, profitability and environment embedded in the supply chain coordination of grocery delivery.
Their findings reveal several dimensions related to functional and non-functional requirements for the effective performance of green crowdsourcing apps for balancing people, environment, and profitability. According to Riedl and Woolley (2016) adequate balancing between the people, profitability and the environment is critical towards the improvement of the quality of life and sustainability.

Transportation apps such as Uber, a mobile app that enables passengers to register as users and motorists to register as transport providers, and Bolt, for ordering rides that are quick, safe and affordable, with grocery are examples of green crowdsourcing apps that have improved the economy (Mwaura, 2016), and created jobs (Taleb et al., 2015) in the countries they operate. These apps have integrated the critical aspects of profitability and people without appropriately balancing the critical triple bottom line of people, profit, and the environment. The potential impact of mobile app innovation for sustainable logistics is critical towards bridging the gap of balancing the critical requirements for meeting people's needs and profitability for the business enterprises and transporters together with contributing to environmental protection. The environmental perspective of the Dilivari App helps strengthen apps like Uber and Bolt that mainly focus on profitability and people. The mobile app innovation not only addresses SDG 13 but also offers fresh insights on how academia can transform research into innovation that helps to shape the supply chain.

METHODS

Design Science Research Approach involves a problem-solving paradigm with the objective of enhancing human knowledge by creating innovative artefacts (vom Brocke et al., 2020; Nunamaker et al., 2017). Figure 1 is a diagram that depicts the research approach that was undertaken.

Figure 1: Diagram illustrating the Decision Science Methodology



The approach follows a sequence of understanding problem relevance, design search process, design of the artefact, design evaluation, research rigour, research contribution and communication. The app was designed in accordance with guidelines for designing innovative technological artefacts for broadening the capabilities of humans and organisations alike (Nunamaker et al., 2017). Literature provides a class of development methodologies that follows many approaches including, but not limited to, waterfall (systematic), iterative, spiral (lifecycle oriented), v-shaped (controlled focused), and agile (highly adaptive) (Kramer, 2018; Afif et al., 2020; Almeida & Simões, 2019).
This study commenced with the waterfall development approach for the system design and implementation of the green crowdsourcing mobile app, Dilivari. An overview of the waterfall model shows that it is advantageous compared to other software development models in terms of producing quality because of the predefined structural flow (Adenowo & Adenowo, 2017). Furthermore, the same structure gives visibility to all stakeholders of the progress of the project as it moves along different steps of the waterfall model (Gorrod & Gorrod, 2004). Mixed evidence illustrates that the waterfall model has shortcomings that must be considered before undertaking to use the model. The highlighted drawbacks and limitations are that it is costly to revert to a surpassed phase and changes in one area impact the next phases due to their dependencies (Dubey et al., 2015). These disadvantages have a significant impact on mobile app development

where the turnaround time is expected to be short and the changes quick and seamless (Flora et al., 2014). Hence the study pivoted to the agile approach that follows an adaptive approach through the development cycle while waterfall is a plan-driven approach (Adenowo & Adenowo, 2017). The basis for incorporating a more agile approach was to enable flexibility, adaptability and shortened time to a minimum deliverable output to compensate for waterfall shortfalls.

Problem relevance

The research study began with an assessment of problem relevance in view of Bormane and Bērziša's (2017) offering a road map of problem relevance for concept design from research to artefact. The problem design involved the engagement of stakeholders to agree on the requirements that informed technological features: system, functional, user and environmental requirements for balancing people, profitability and environment was a major part of the green crowdsourcing app, as discussed in detail by previous studies (Nieman & Bennett, 2002; Fahey & Narayanan,1986). System requirements were developed to facilitate designing software meeting the requirements of users and involved documentation of system specifications (Akanmu et al, 2013). Several scholars argue the need for the development of requirements: Functional requirements (Bahill & Madn, 2017), user requirements (Vijiyalakshmi & Sivaraj, 2016), and environmental requirements (Mkansi, 2013), were gathered, analysed, documented, and translated into functional and non-functional requirements to effectively cater for the design as per table 1.

Table 1: Requirements gathered

Requirements	Features
Functional	Features for effective functioning of a crowdsourcing device were user account, log-in ability, request for delivery, view available delivery request, view available delivery requests, view receipts, pick-up delivery, address completion, distance calculation for pick-up and sign delivery on receipt.
User requirements	The user requirements create a profile and sign into the app, request collection and parcel delivery, and view a list of delivery requests and make an offer to deliver a package en route to their destination.
Environment requirement	The environment requirements are a smartphone powered by either an Android or Apple iOS operating system, internet connection and an active email account. The user must have access to download the app from any of the online app stores such as Google Play, Apple App store, Windows store, or marketplace
Actors and roles	There must be an administrator, requesting user, receiving user and the app

Design as a search process

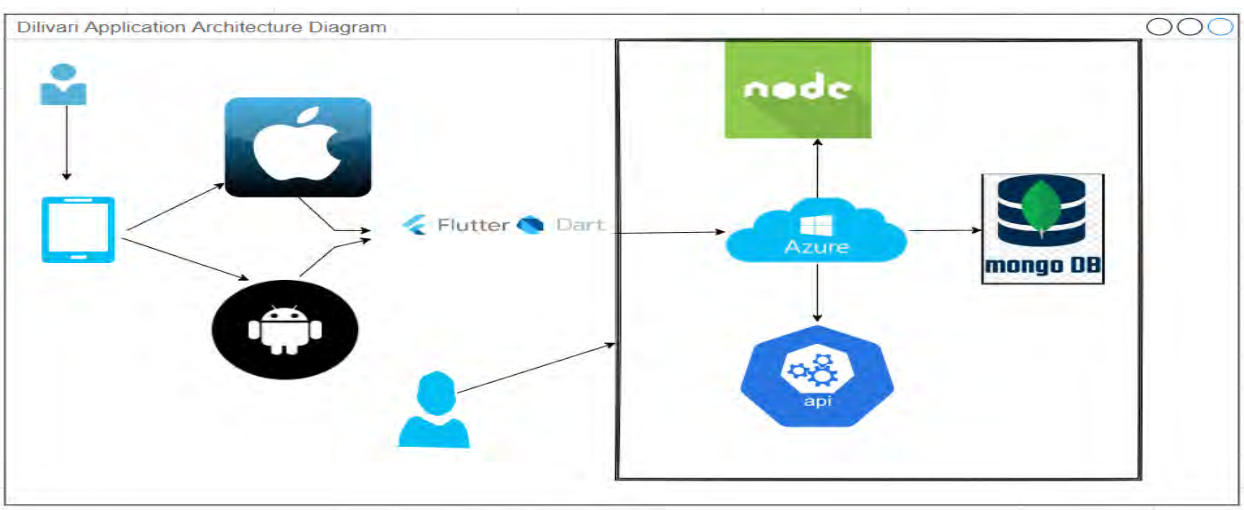
This section provides the design search process that was undertaken during the development of the Dilivari artefact. The approach involved designing the app architecture, design diagrams, use case diagrams and flow chart diagrams which are key for the implementation step.

<u>Application architecture</u>

App architecture provides specifications for the implementation of an information system (Whitten, 2000). The Dilivari app architecture was built to meet the requirements specified by the stakeholders. A multi-layered diagram was developed providing the user interface, user communication layer, information retrieval layer and database layer to provide a blueprint of the system that defined the structure of the system. The development process adhered to major principles for the implementation of architecture, i.e., separating concerns, minimising interdependencies, and isolating fundamental components for the modularisation of the app into smaller functions (Costa et al., 2020). The architecture for the Dilivari app was developed with three components, namely the user interface (UI) which is the screens of the mobile app, the data model layer that defined the data structure, constraints and API calls to the final component, and the database containing the data.

Use case diagrams were used to capture system requirements that included internal influences and external influences, covering mainly design requirements (Mule & Waykar, 2015). The use case diagram was used as a tool to give a visual depiction of the possible interactions between the user and the Green Dilivari app. Class diagrams is another form of modelling in software development. A class diagram was utilised to visualise the granular levels of an app by explaining the different components of the app. Thus class diagrams are at the level of describing the app classes with their attributes and operations (Bahill & Madni, 2017). An entity diagram was specifically used for modelling the data layout for the database structures for storing information. An entity-relationship diagram is a graphical representation used in the design phase showing the relationships between individuals, things, concepts, or events in an information system (Boston University Center for Teaching & Learning, n. d). The developed architectural diagram had a user interface, a user communication layer, and an information retrieval layer. The user interface was critical in showing the mobile device screen and the communication layer harboured the code running on iOS or Android responding to the instructions as the user makes commands on the app. Finally, the information retrieval layer, the Flutter code, communicated with MongoDB by making API calls and was the database containing all user information. This is shown in Figure 2**.**

Figure 2: Dilivari mobile app high level.

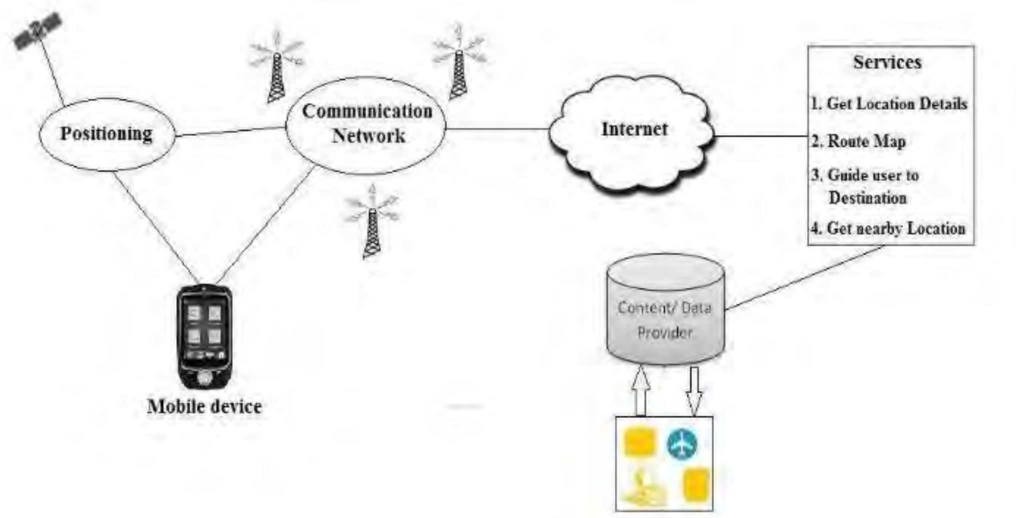


Components for location

The crowdsourcing app had to support location-based services (LBS) to help in the determination of a user's geographical location and provide useful information to enhance user experience

(Huang et al. 2021). The service of maps and location is rendered by Google Maps and the API service assisted with app functions that require a map's information (Jha & Chourasia 2011). This is shown in Figure 3.

Figure 3: Required components for obtaining location (Jha & Chourasia 2011)



The Google Maps API is a service from Google that enables an app to gain access to functions such as locations on a map (Zola, 2022), identifying different routes between location points and estimating travel time by different modes of travelling such as a train, walking, a car or cycling. The service generates an API key that uniquely identifies the app making a call to Google. The API key is also used for billing; high-traffic calls from the app result in Google requiring payment for the calls (Juviler, 2023). The location was then listed as a challenge because it was dependent on several factors to report accurate results. These factors include: a user must be in an area with good connectivity and reception, have a data connection or Wi-Fi and the location must already be listed on Google Maps.

Design as an artefact

Following design as a search process is design as an artefact which includes implementation, coding, development, and prototyping. The implementation discusses the technologies and coding used to implement green crowdsourcing mobile apps.

Implementation

A well implemented app impacts positively on user satisfaction and motivation to continue using the app (Samra et al., 2020). For the implementation of the screens and navigations Flutter was the tool used for development. Flutter is a software development kit used for creating hybrid apps, i.e., mobile apps that can be installed on both Android and iOS operating systems while having one code base. Flutter is best for creating high-performance platform-independent apps (Ameen & Mohammed, 2022). Furthermore, with limited previous studies, Flutter proved favourable in development speed and budget reduction. There were many kinds of apps from which to draw comparison. In this case comparison was drawn between native and hybrid apps because they had fewer disadvantages to the other types. A summary of previous literature suggest that native apps have a better memory consumption (Wasilewski & Zabierowski, 2021), CPU utilisation and a smaller package bundling. Further comparisons between Flutter and native apps are greatly discussed by Hussain et al. (2021).

Visual Studio Code was used to write the code, compile and test it. It was also used for the process of packaging the app for deployment to iOS and Android platforms. A summary from previous literature suggests that Visual Studio Code is disadvantaged because it heavily consumes memory compared to an integrated development environment (IDE) such as Android Studio and IntelliJ. Furthermore, it does not have an intuitive user interface to configure optimal functionality. In contrast, being an open source, IDE was found to be advantageous and fast, powerful, customisable as could be extended with plugins for developing, simulation and testing (Zola, 2022). Another IDE that was used was Xcode, which is an IDE for MacOS. However, it is only limited to Apple devices (Fojtik, 2020), but great for testing and has an instinctive interface for first-time users. Xcode was mainly used for testing the app for iOS deployment and packaging the deployment bundle.

Database

Sahatqija et al. (2018) guided the developers in choosing MongoDB for the green crowdsourcing app on the basis of scalability, flexibility, performance, query language, security, data replication, licensing and availability. The open-source database without schema, faster in comparison to SQL databases, does not require the creation of a structure for the data to be inserted but simply needs a JSON object to do an insert. MongoDB was hosted in Azure, which is a Microsoft cloud computing service for the purpose of app management using Microsoft data centres. It supports different technologies, programming languages, tools, frameworks and also offers software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS) (Microsoft Azure 2020). The main function for Azure is hosting MongoBD for storing the green crowdsourcing mobile app Dilivari's data. There are many other cloud computing service providers. Azure was however selected because of its affordability and full security control as compared to other well-established cloud service platform like Amazon Web Services (AWS).

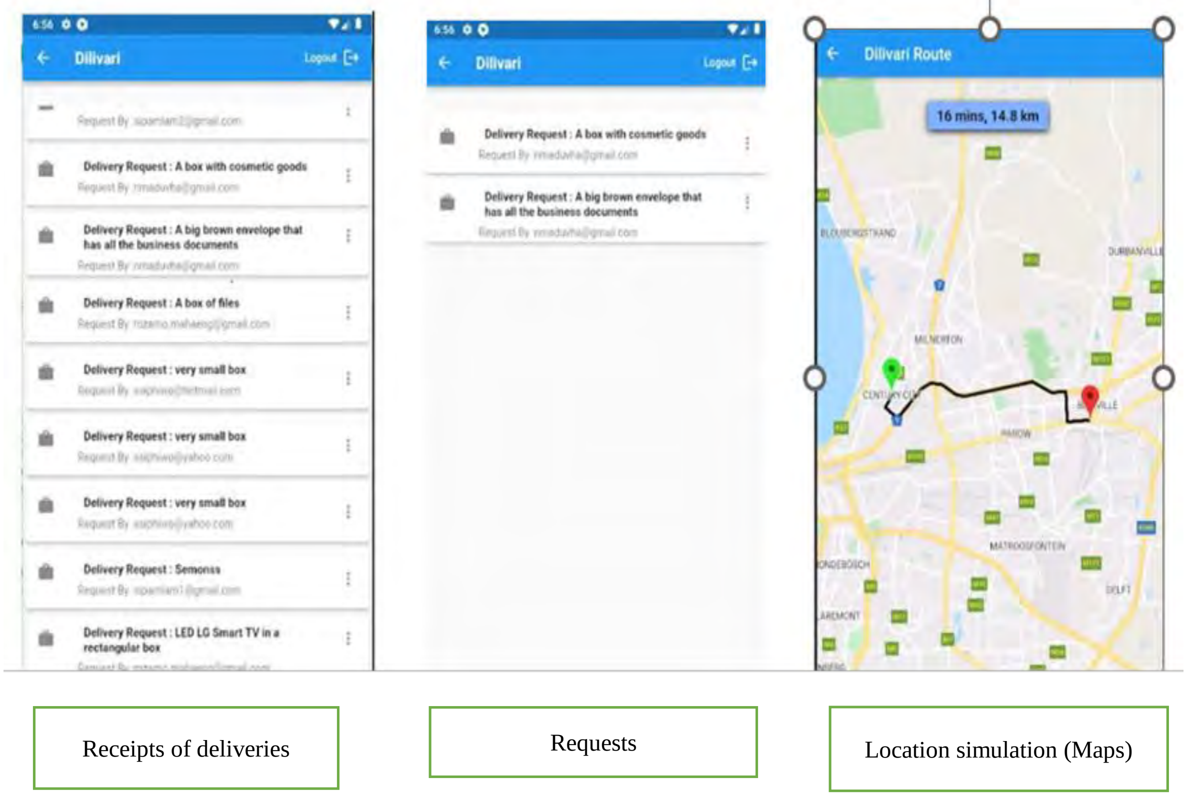
Design evaluation and rigour (testing)

Design and evaluation involved conducting a quality assurance mechanism for ensuring the artefact meets the requirements. The evaluation that was conducted undertook functional and compatibility testing, performance testing, middle layer testing and confidentiality layer testing.

Functional and compatibility testing

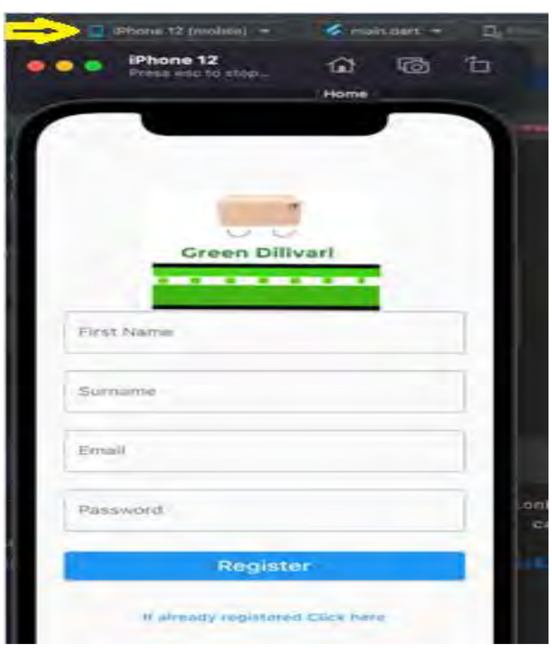
Preliminary testing was conducted in accordance with the principles for the development of crowdsourcing software apps. Akanmu et al. (2013) argue that testing apps is critical to meet the requirements and evolving technology and customer expectations in terms of aesthetics and divergent expectations. Preliminary testing included test automation, location simulation, engaging with third parties, understanding physical characteristics, user experience, connectivity-related testing, dealing with fragmentation, end-to-end integration testing, performance and security (Knott 2015). The app was tested for both functional and compatibility testing. Arif and Ali (2019) argue that functional testing on mobile apps is to validate the actual functionality of the app. Figure 4 provides a functional map for functional testing.

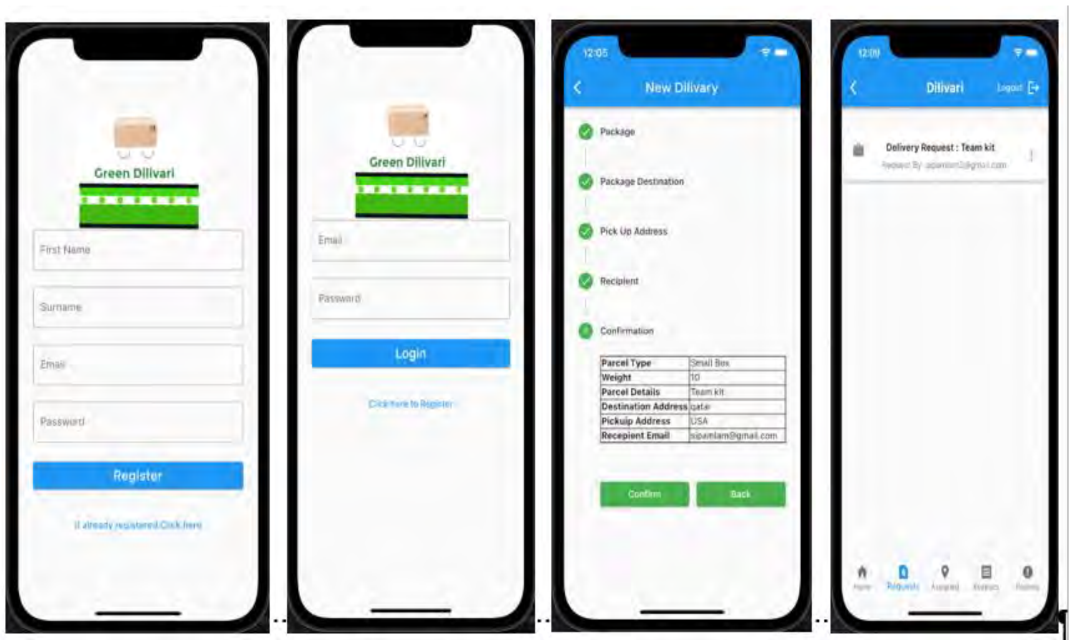
Figure 4: Functional Testing –Map Showing Receipts, Deliveries, Requests and Location



In accordance with the principles for designing crowdsourcing software compatibility testing was conducted for the Dilivari Application. Compatibility testing involves ascertaining whether the application is suitable to be used with the different phone models (Wu et al., 2013; Stol, and Fitzgerald, 2014). The results for compatibility testing are provided in Figure 5 below.

Figure 5: Simulator Testing





<u>Performance testing</u>

Performance testing is used for validating an app's architecture. It evaluates all layers of the app from the front-end, the middle layer and the back-end of the technology stack (Knott, 2015). The Dilivari app architecture was defined as having the following components: a user interface (UI) that is the screen of the mobile app and a data model layer that defines the data structure, constraints and API calls to the final component, and the database. On testing performance, testing was conducted in units and then integrated to validate the seamless integration of the architecture components. The screens below are a sample of the testing done on the UI:

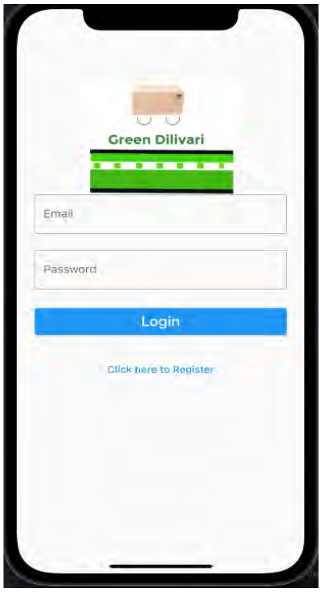
Figure 6: Sample testing on UI

Sign-up Screen

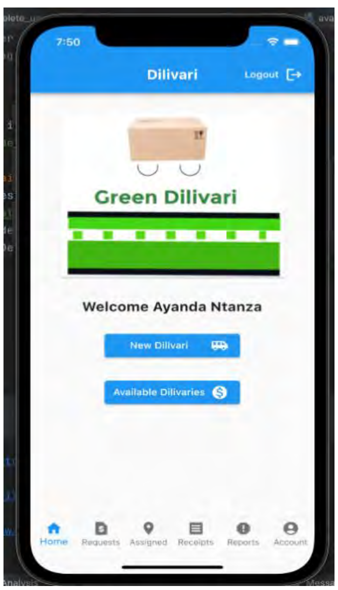
Testing was done to check that the user was able to write, delete and edit in the text field as well as toggle betwee text fields.

Testing was done on the 'Register' button to chec whether it was clickable and able to navigate to the correct screen once user inputs were added.

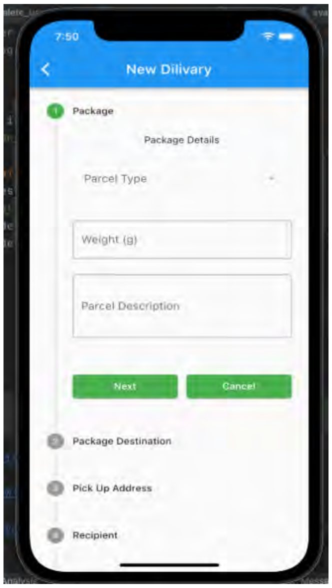
Below the register is a clickable link that is supposed to land on the login; it was also tested.



Login Screen
Testing was done to check that the user was able to add inputs on the email and password text boxes, and able to navigate between fields.
Testing was also done to check that the 'Login' button is clickable and led to the home page if the email-passw combination was correct
Testing was also performed to verify that the 'Click here to Register' button was clickable and navigated to the 'Register' page.



Home Screen
Testing was done to verify that the correct user's name ID displayed upon login.
It was done to check that the two buttons to create new delivery requests and the button to check already available deliveries were clickable and navigated to the correct screen.
Testing on the footer buttons confirmed that they were all clickable and navigated to appropriate screens.
Testing verified the 'Logout' navigation button at the top terminated the sessions and navigated back to the logi screen.

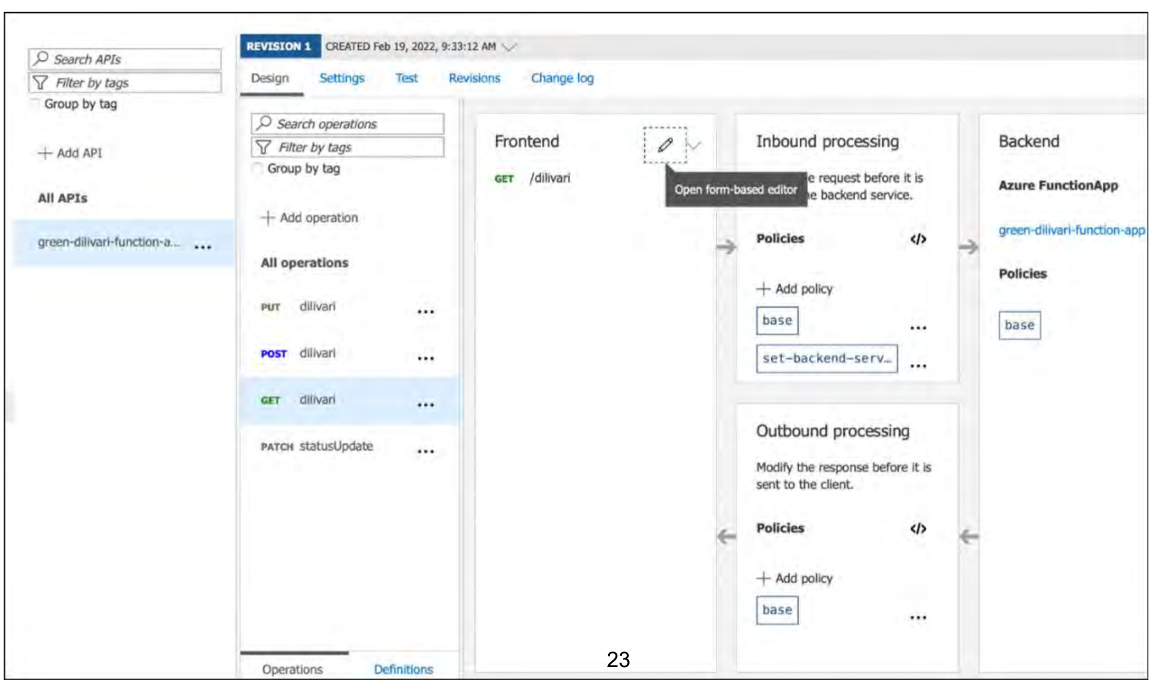


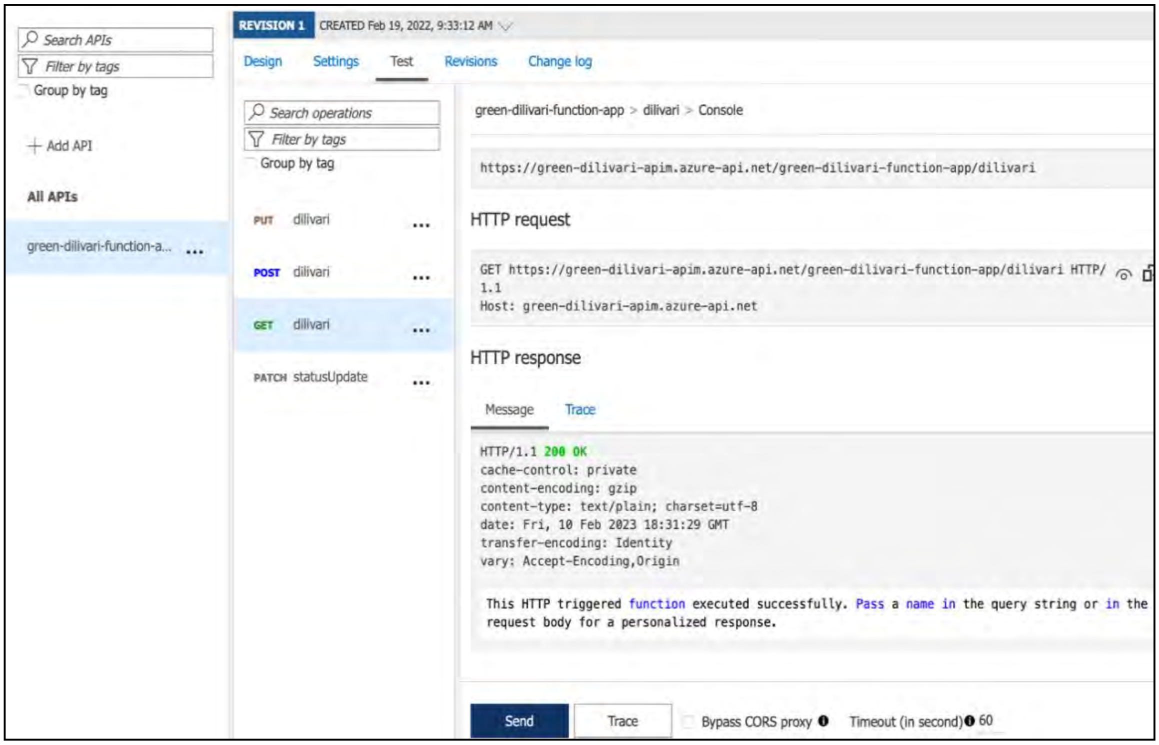
New Delivery Screen
Testing checked that the flow worked, and the navigation button helped move to the next and previou steps.
Testing verified that upon entering an address a dropdown list of suggestions appeared as an auto-complete to the address.
Testing also vetted that the 'Confirm' button submitted the delivery request.

A similar approach of testing was completed for the rest of the screens verifying that the sequence of actions performed by the user yielded the expected response from the UI.

Middle layer testing
The middle layer is a link between the UI and the back end – the database. This layer exposed the REST API that front end uses to interact with the database. This layer accepted HTTP requests from the UI and utilised CRUD (create, read, update, delete) operations connecting to the database to create data and perform data manipulations and to query and get results, while also processing with those results. The middle layer was spread across the code and Microsoft Azure. The REST API endpoint and operations were housed in a Microsoft Azure function app that has the functionality to enable API management. The starting point was to test the REST API as a separate component by testing the API operations. Following Microsoft Azure, the testing was conducted for making queries and retrieving data: We used the GET operation for maing queries and retrieving data. The GET operation is illustrated in figure 7.

Figure 7: GET testing on Microsoft Azure

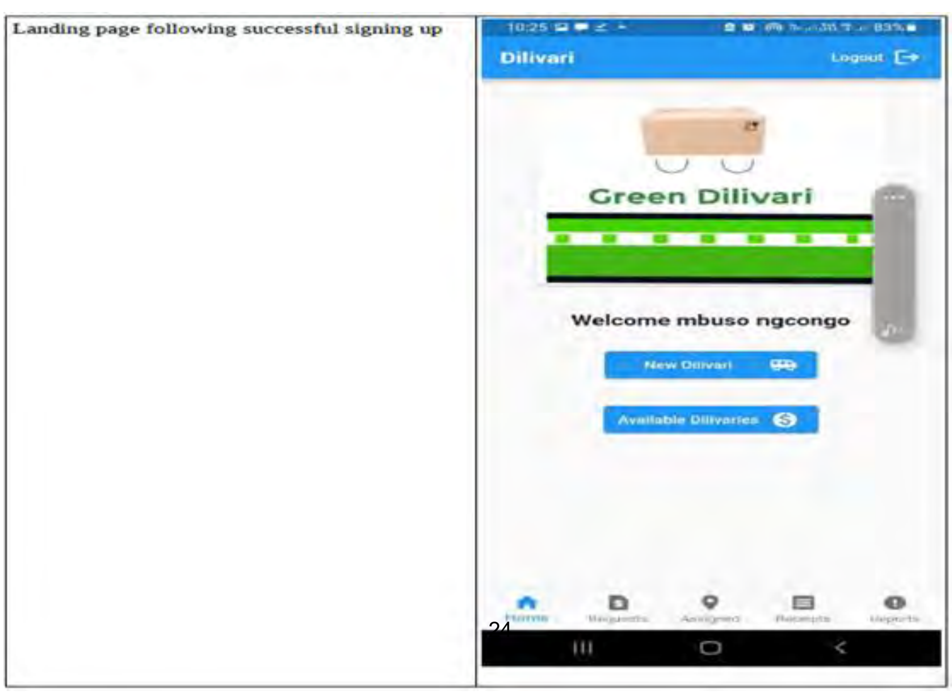




<u>Confidentiality security testing</u>

Security testing is a significant part of application testing; validating security measures uncovers critical security gaps including threats. Paul and Aithal (2019) emphasises the importance of security testing highlighting the challenge imposed by hackers, most of who are unethical. Many studies point to about six general security aspect that must should be tested (Ojenge et al., 2014); these are integrity, confidentiality, non-repudiation , authentication, authorisation, availability. For security testing only, the log in function was tested to ensure that only registered users could gain access, authentication testing. Figure 8 below illustrates the apps landing page after a user has successfully logged in.

Figure 8: Testing of the log in functionality



RESULTS

As this study is based on development, the findings are the actual mobile app developed, which serves as the extension of the design methods that cover design evaluation (testing and observation of the artefact) and research rigour (accuracy and precision in line with research). The study undertook the development of a mobile app artefact, used for the delivery of groceries across different areas to meet the needs of the different stakeholders in urban and rural areas. The App is accessible on Google Play and Apple App store under the title Dilivari. Some of the core features that were deduced through problem relevance, design as a process, and design artefact using a waterfall methodology produced the unified User Interface (UI). The app is encapsulated. This is accomplished by hosting the Dilivari mobile app in the Microsoft Azure shell flyout. According to Wulf & Blohm (2020) the unified User Interface (UI) uses responsive web design principles to provide an optimal viewing and interaction experience for any screen size, device, or orientation. The Dilivari app went through design and evaluation as a quality assurance mechanism for testing the artefact before deployment for use by the stakeholders. Design and evaluation involved conducting the quality assurance mechanism to ensure the artefact meets the requirements. The evaluation that was conducted undertook functional and compatibility testing, performance testing, middle layer testing and confidentiality layer testing.

Functional testing was conducted to determine the extent to which the app met requirements, evolving technology and customer expectations (Akanmu et al, 2013). Knott (2015) highlights that performance testing used for evaluating all layers of the App from the front-end, the middle layer to the back end of the technology stack is required. Middle layer testing was conducted to determine whether the middle layer linked very well between the UI and the back end – the database. Finally, Paul and Aithal (2019) emphasises the importance of security testing highlighting the challenge imposed by hackers, most of whom are unethical. The App performed to optimal level across computer experimentation. The App is undergoing a grocery pilot study in Pretoria to determine its suitability for the purpose of meeting requirements for the grocery delivery, people, and profitability on the basis of the specifications determined.

DISCUSSION AND CONCLUSIONS

The current study adopted a design science approach for the development of the green Dilivari crowdsourcing App. The green crowdsourcing app development was unique as opposed to the traditional outsourcing strategies that are characterised by contracts between two parties. The current study extended the findings of Sipamla and Mkansi (2019) by developing a mobile app, which was the main research objective of the study. It completes the last two stages of design methods by outlining the research contribution (application to theory, practice, and methods), and lastly communication (publication of the app and research). The development of the Dilivari app has clear implications for task decomposition, coordination, and communication, planning and scheduling, quality assurance, knowledge and intellectual property, motivation, and remuneration. Furthermore, the app has critical implications in the context of developing and emerging economies.

Practical contribution

In practice, the innovation benefits people, profit, and the environment's sustainability. First, the app provides an opportunity for the unemployed to deliver groceries, providing an incentive for transportation of groceries during the peak hours and building on their own flexible work. The Dilivari App has improved the processes of coordinating the supply and distribution of groceries for urban and rural population effectively and efficiently.

Secondly, the app helps grocery retailers address the major issue of delivering low product volumes that increase the costs of delivering groceries to consumers. The app allows voluntary delivery by consumers that have extra space to earn additional income. The app significantly

reduces costs of grocery retailers and improving profit margins. The green Dilivari app innovation technology is increasingly enabling and driving the transformation of these expensive services. The affordability of the Dilivari App technology makes it convenient for use by independent shoppers, transporters, and couriers in developing economies in Africa.
Thirdly, the Dilivari App reduces emissions and congestion systematically addressing increasing emissions. The app optimises the logistics architecture as grocery deliveries use the most appropriate mode of transportation, load capacities, and effective routing substantially reducing carbon emissions. Furthermore, the app provides an opportunity for the delivery of groceries by taking advantage of passengers' spare capacity on already completed journeys, contributing positively to economic, social, and environmental wellbeing. The drivers use the unused cargo capacity to deliver groceries thereby reducing traffic congestion as drivers with extra capacity support in transportation of groceries every day. The app reduces congestion and emissions impacting the environment as highlighted by Batool et al., (2022), Chatti (2021), Haini (2021), Kwakwa et al., (2023), Xie et al., (2017), and Zaman et al., (2017).

Methodological contribution

The development of the Dilivari app provides a blueprint for the development of the app as it highlights mechanisms for the development of crowdsourcing apps in emerging economies. The study also provides an appropriate guide for converting research findings into software design. Systematically, it extends concepts for development of crowdsourcing apps to developing an app for balancing people, profit, and environmental sustainability. Furthermore, it strengthens the concepts of designing artefacts for problem solving paradigms for improvement of the human livelihoods by creating innovative artefacts (vom Brocke et al. 2020).

Theoretical contribution

Traffic congestion, environmental management, climate change and poverty alleviation have continued to be major challenges in emerging economies. The app provides mechanisms for contributing towards the reduction of traffic congestion, profitability and effective environmental management which is impacting Africa. The congestion because of stakeholders transporting groceries has continued to contribute to 21% of global emissions (Ritchie, 2020). Furthermore, Africa has continued to contribute to global CO_2 emissions (DeWitt et al., 2019). Using the app supports the reduction of these emissions as traffic already on the road is optimised to distribute a few groceries, reducing distance, time, and congestion. This is critical in environmental management, profitability, and alleviation of poverty.

REFERENCES

Adenowo, A. A., & Adenowo, B. A. (2013). Software engineering methodologies: a review of the waterfall model and object-oriented approach. *International Journal of Scientific & Engineering Research*, 4(7), 427-434.

Afif, A.N., Noviyanto, F., Sunardi, S., Akbar, S.A. and Aribowo, E. (2020). Integrated application for automatic schedule-based distribution and monitoring of irrigation by applying the waterfall model process. *Bulletin of Electrical Engineering and Informatics*, 9(1), 420-426.

African Union. (n. d). *Goals and Priority Areas of Agenda 2063*. https://au.int/en/agenda2063

Akanmu, A., Anumba, C., & Messner, J. (2013). Scenarios for cyber-physical systems integration in construction. *Journal of Information Technology in Construction*, 18(12), 240-260.

Almeida, F., & Simões, J. (2019). Moving from waterfall to agile: Perspectives from IT Portuguese companies. International *Journal of Service Science, Management, Engineering, and Technology (IJSSMET)*, 10(1), 30-43.

Ameen, S. Y., & Mohammed, D. Y. (2022). Developing cross-platform library using flutter. *European Journal of Engineering and Technology Research*, 7(2), 18-21.

Arif, K. S., & Ali, U. (2019, January). Mobile Application testing tools and their challenges: A comparative study. *2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-6). IEEE.

Bahill, A. T., & Madni, A. M. (2017). *Tradeoff decisions in system design*. Springer International Publishing.

Bassi, H., Lee, C. J., Misener, L., and Johnson, A. M. (2020). Exploring the characteristics of crowdsourcing: An online observational study. *Journal of Information Sciences*, 46(3), 291–312. https://doi.org/10.1177/0165551519828626

Batool Z, Raza SMF, Ali S, Abidin SZU (2022). ICT, renewable energy, financial development, and CO2 emissions in developing countries of East and South Asia. *Environment Science Pollution Research*, 1-11. https://link.springer.com/article/10.1007/s11356-022-18664-7.

Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.M., Conradsen, M.L., Di Marco, D., Distel, A.P., Dobusch, L., Dorler, ¨ D., Effert, A., Fecher, B., Filiou, D., Frederiksen, L., Gillier, T., Grimpe, C., Gruber, M., Haeussler, C., Heigl, F., Hoisl ... & Xu, S.M. (2022). The open innovation in science research field: a collaborative conceptualisation approach. *Industrial Innovation*, 29 (2), 1–50.

Bormane, L. and Bērziša, S. (2017). “Role of ‘bridge person’ in software development projects”. *In International Conference on Information and Software Technologies* (pp. 3-14*).* Springer, Cham. Druskininkai, Lithuania, October 12–14, 2017.

Boston University Center for Teaching & Learning (n.d.). *Experiential Learning.* The Centre for Teaching and learning. https://www.bu.edu/ctl/guides/experiential-learning/

Chatti, W. (2021). Moving towards environmental sustainability: information and communication technology (ICT), freight transport, and CO2 emissions. *Heliyon*, 7(10). https://www.cell.com/heliyon/pdf/S2405-8440(21)02293-3.pdf

Costa, B., Pires, P. F., & Delicato, F. C. (2020). Towards the adoption of OMG standards in the development of SOA-based IoT systems. *Journal of Systems and Software*, 169, 1-29. https://doi.org/10.1016/j.jss.2020.110720

DeWitt. J.C., Blossom. S.J., and Schaider. L.A. (2019). Exposure to per-fluoroalkyl and polyfluoroalkyl substances leads to immunotoxicity: Epidemiological and toxicological evidence. *Journal Exposure Science Environmental Epidemiology*, 29(2), 148–156.

Dubey, A., Jain, A., & Mantri, A. (2015). Comparative study: waterfall v/s agile model. *International Journal of engineering sciences & research technology*, 4(3), 70-75.

Edgar, J., Murphy, J., & Keating, M. (2016). Comparing traditional and crowdsourcing methods for pretesting survey questions. *Sage Open*, 6(4). https://doi.org/10.1177/2158244016671770

Fahey, L. and Narayanan, V.K. (1986). *Macro Environmental Analysis for Strategic Management* (The West Series in Strategic Management). St. Paul, Minnesota: West Publishing Company.

Flora, H. K., Wang, X., & Chande, S. V. (2014). An investigation into mobile application development processes: Challenges and best practices. *International Journal of Modern Education and Computer Science*, 6(6), 1-9.

Fojtik, R. (2020). Swift a New Programming Language for Development and Education. In: Antipova, T., Rocha, Á. (Eds.), *Digital Science 2019* (pp. 284-295). Springer, Cham. https://doi.org/10.1007/978-3-030-37737-3_26

Franzoni, C., Poetz, M., & Sauermann, H. (2022). Crowds, citizens, and science: a multi-dimensional framework and agenda for future research. *Industry and Innovation*, 29(2), 251-284.

Gorrod, M., & Gorrod, M. (2004). The software development lifecycle. *Risk Management Systems: Process, Technology and Trends*, 93-120.

Haini, H. (2021). Examining the impact of ICT, human capital and carbon emissions: evidence from the ASEAN economies. *International Economics*, 166(1),116–125. https://doi.org/10.1016/j.inteco.2021.03.003

Hedges, M., & Dunn, S. (2017). *Academic crowdsourcing in the humanities: Crowds, communities and co-production*. Chandos Publishing.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1):75–105.

Huang, H., Yao, X. A., Krisp, J. M., & Jiang, B. (2021). Analytics of location-based big data for smart cities: Opportunities, challenges, and future directions. *Computers, Environment and Urban Systems*, 90, 101712. https://doi.org/10.1016/j.compenvurbsys.2021.101712

Hussain, H., Khan, K., Farooqui, F., Arain, Q. A., & Siddiqui, I. F. (2021). Comparative Study of Android Native and Flutter App Development. *Memory*, 47, 99-102. https://www.researchgate.net/profile/Hina-Hussain-2/publication/361208165_Comparative_Study_of_Android_Native_and_Flutter_App_Development/links/635a857c12cbac6a3efff6b6/Comparative-Study-of-Android-Native-and-Flutter-App-Development.pdf

Jha, A. N., & Chourasia, R. (2014). Location Based Services in Android with Google maps integration. https://d1wqtxts1xzle7.cloudfront.net/35703258/Location_Based_Services_in_Android_-libre.pdf?1416814571=&response-content-disposition=inline%3B+filename%3DLocation_Based_Services_in_Android_with.pdf&Expires=1693209450&Signature=QiHMCcskd6w7opbmquSO5s4evYe5IDMMRt8QnhMFOzcDezVjoClKrxOhBevDPqlPdIzQBHmBmYJsz32eFsDkNNqrclRLqESNvRpuTl2NsoWYPlmP3SkAxYlMkFBGZbJ6EW2q-kEwRGo30-Cm5j5QWazktdi2asLzzIb8b1DLT4W0oRaIFls4JivBWqV04huyUiKANPfMg4yGwHedtKMRpAQ

QyjeYpahrf1j5CPsmL4P-yCZ0s4x4oMOukK5YgYkREAvxVW6KJ9GwDbdJEIigB27wQZ5Y464eHHNkaVLF2vzRa8uDAjVFZLoOMMvWHmTEr2WHWr7wDbdhDn9enVvTCQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Juviler, J. (2023 August 25th). What is an API Key? (And Are They Secure?).HubSpot Blog | What is an API Key? (And Are They Secure?) (hubspot.com)

Knott, D. (2015). *Hands-on mobile app testing: a guide for mobile testers and anyone involved in the mobile app business*. Addison-Wesley Professional.

Kramer, M. 2018. Best practices in systems development lifecycle: An analyses based on the waterfall model. *Review of Business and Finance Studies*, 9(1), 77-84.

Kwakwa, P. A., Adjei-Mantey, K., & Adusah-Poku, F. (2023). The effect of transport services and ICTs on carbon dioxide emissions in South Africa. *Environmental Science and Pollution Research*, 30(4), 10457-10468.

Law, E., Gajos, K. Z., Wiggins, A., Gray, M. L., & Williams, A. (2017, February). Crowdsourcing as a tool for research: Implications of uncertainty. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1544-1561). https://doi.org/10.1145/2998181.2998197

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1), 1-23.

Mkansi, M. (2013) *Management of the e-retail supply distribution network: A resource-based view of the UK grocery Sector* [Doctoral dissertation, Bolton University].

Mule, S. S., Waykar, Y. (2015). Role of USE CASE diagram in s/w development. *International Journal of Management and Economics*.

Mwaura, G (2016, March 18). *Uber: The beauty and wrath of disruptive technologies in Africa. The New Times.* https://www.newtimes.co.rw/article/128159/Opinions/uber-the-beauty-and-wrath-of-disruptive-technologies-in-africa

Nidhra, S. and Dondeti, J. (2012). Black box and white box testing techniques-a literature review. *International Journal of Embedded Systems and Applications*, 2(2), 29-50.

Nieman, G. and Bennett, A. 2002. *Business management a value chain approach*. Van Schaik Publishers.

Nunamaker, J.F., Twyman, N.W., Giboney, J.S. and Briggs, R.O. (2017). Creating high-value real-world impact through systematic programs of research. *MIS Quarterly*, 41(2) 335-351.

Ojenge, W., Okelo-Odongo, W., and Ogao, P. (2014). Use of GA-Optimized NN to Predict DVB-T2 Receiver Spectrum Holes to Accommodate Burden GSM Voice Calls. *International Journal of Simulation- Systems, Science and Technology* 16(4):121-129

Paul, P., & Aithal, P. S. (2019). Mobile applications security: An overview and current trend. *Proceedings of National Conference on Research in Higher Education, Learning and Administration*, 1(1), 112-121.

Riedl, C., & Woolley, A. W. (2017). Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. *Academy of Management Discoveries*, 3(4), 382-403.

Ritchie, H. (2020). You want to reduce the carbon footprint of your food? Focus on what you eat, not whether your food is local. Our world in data, 24. https://climatecommentary.com/2021/01/23/you-want-to-reduce-the-carbon-footprint-of-your-food-focus-on-what-you-eat-not-whether-your-food-is-local/

Sahatqija, K., Ajdari, J., Zenuni, X., Raufi, B., & Ismaili, F. (2018, May). Comparison between relational and NOSQL databases. In *2018 41st international convention on information and communication technology, electronics and microelectronics* (pp 0216-0221). IEEE. https://www.researchgate.net/profile/Jaumin-Ajdari-2/publication/326699854_Comparison_between_relational_and_NOSQL_databases/links/5b805b2fa6fdcc5f8b647c74/Comparison-between-relational-and-NOSQL-databases.pdf

Samra, B., Richard-Carpentier, G., Kadia, T.M., Ravandi, F., Daver, N., DiNardo, C.D., et al., (2020). Characteristics and outcomes of patients with therapy-related acute myeloid leukemia with normal karyotype. *Blood Cancer Journal*,10 (5), 47-52

Sheehan, K. B. (2018). Crowdsourcing research: Data collection with Amazon's Mechanical Turk. *Communication Monographs*, 85(1), 140–156.

Stol, K. J., & Fitzgerald, B. (2014, May). Two's company, three's a crowd: a case study of crowdsourcing software development. *In Proceedings of the 36th International Conference on Software Engineering* (pp. 187-198). Association for computing machinary. https://dl.acm.org/doi/10.1145/2568225.2568249

Taleb, Z., Ahmadi, A., & Musavi, M. (2015). The effect of m-learning on mathematics learning. *Procedia-Social and Behavioral Sciences*, 171, 83-89. https://doi.org/10.1016/j.sbspro.2015.01.092

Vijiyalakshmi, K. and Sivaraj, D. (2016) Synergistic antibacterial activity of barium doped TiO2 nanoclusters synthesized by microwave processing. *Royal society of Chemistry Advances*, 1(6), 9663-9671.

Vom Brocke, J., Hevner, A., & Maedche, A. (2020). Introduction to design science research. In J. Vom Brocke, A. Hevner & A. Maedche (Eds.), *Design science research. Cases* (pp.1-13). Springer.

Wulf, J. and Blohm, I. (2020). Fostering Value Creation with Digital Platforms: A Unified Theory of the Application Programming Interface Design. *Journal of Management Information Systems***,** 7 (5), 694–709.

Wu, W. Tsai, W. T. and Wei. L. (2013). An Evaluation Framework for Software Crowdsourcing. *Frontiers of Computer Science***,** 7 (5), 694–709.

Xie R, Fang J, Liu C (2017). The effects of transportation infrastructure on urban carbon emissions. *Applied Energy*, 196, 199–207. https://doi.org/10.1016/j.apenergy.2017.01.020

Wasilewski, K., & Zabierowski, W. (2021). A comparison of java, flutter and kotlin/native technologies for sensor data-driven applications*. Sensors*, 21(10),1-16.

Wu, W. Tsai W. T. and Wei. L (2013). Creative software crowdsourcing: from components and algorithm development to project concept formations. *International Journal of Creative Computing*, 1 (1): 57- 57-91.

Whitten, J.L. (2000). Application, architecture and modelling. In L.D., Bentley, K.C., Dittman & J.L., Whitten (Eds.), *System analysis and design methods*. Irwin/McGraw Hill. https://cs.nyu.edu/~jcf/classes/g22.2440-001_sp05/slides/session6/g22_2440_001_c64.pdf

Zaman K, Moemen M.A.E, and Islam T (2017) Dynamic linkages between tourism transportation expenditures, carbon dioxide emission, energy consumption and growth factors: Evidence from the transition economies. *Current Issue Tour*, 20(16), 1720–1735.

Zola, A. (2022 September, 8th). Google Maps.Techtarget. What is Google Maps and how do you use it? (techtarget.com).

DECISION SCIENCES INSTITUTE

A Methodology for GPS Spoofing Detection and Mitigation Technique

Rakesh K. Sharma
University of Maryland Eastern Shore
rsharma@umes.edu

Tarun Dhar Diwan
Atal Bihari Vajpayee University
tarunctech@gmail.com

ABSTRACT

In this study, the efficiency of various recognition and mitigation approaches for spoofing attacks on Global Positioning System (GPS) are evaluated using an experimental example. Assessing the impact on navigation accuracy in the context of GPS spoofing detection and mitigation strategies is critical to determining the effectiveness of these techniques in preserving accurate placement. Overall, impact on navigational accuracy conclusion emphasizes the vulnerability of GPS systems to spoofing attacks and emphasizes the necessity for dependable and efficient mitigating mechanisms to assure accurate and trustworthy navigation information. The detecting effectiveness analysis showed that both of the machine learning (ML) classification techniques and signal anomaly detection methods worked well in identifying spoofing assaults in diverse contexts.

KEYWORDS: Mitigation Approaches, Spoofing Attacks, Global Positioning System (GPS) Spoofing Detection, Signal Anomaly Detection, ML Classification.

INTRODUCTION

GPS spoofing is the manipulation of IoT devices to create false information, leading to security risks such as unauthorized entry and navigation system manipulation. Research has focused on developing GPS spoofing detection systems that assess signal properties and use ML methods to recognize and counteract such attacks. These systems can be monitored in real-time using specialized hardware and software techniques, such as a GPS signal generator or software-defined radio (SDR). Fake GPS signals can be created by changing location, time, or other details, and monitored to determine their impact on the intended GPS receiver. Signal analysis involves comparing received GPS signals to expected ones, identifying any oddities, irregularities, or deviations that could indicate a spoofing attack. Statistical techniques like outlier detection, anomaly detection, or pattern recognition are used to identify spoofing efforts. Cryptographic methods for GPS signal authentication are also explored, such as digital signatures, message authentication codes (MACs), or cryptographic hashes.

Jha et al. (2020) used ML and artificial intelligence techniques to detect GPS spoofing, training classification or anomaly detection models using labeled datasets of real and fake GPS signals. Dang et al. (2022) investigated mitigation strategies and protective measures to reduce the effects of GPS spoofing attacks. Further testing and evaluation were conducted to validate the proposed detection and mitigation approaches. Field testing or real-world scenarios were used to evaluate the effectiveness, precision, and dependability of the methodologies. False positive and false negative rates of detection algorithms were also examined, and the methodology was recorded. The main conclusions and suggestions for preventing and detecting GPS spoofing attacks are provided, along with recommendations for improving the resiliency and security of GPS systems. This methodology helps build reliable and secure GPS systems.

LITERATURE REVIEW

The study on GPS spoofing detection is still in its early phases, although there has been tremendous improvement in recent years. A variety of excellent spoofing detection approaches have been developed, and these techniques are getting more dependable and inexpensive. As the threat of GPS spoofing grows, it is critical to maintain study on this critical problem. Table 1 shows an overview of the literature:

Table 1: Review of Literature based on GPS spoofing problem

Author(s)	Technique Used	Solution
Aissou et al. (2022)	Instance-based supervised ML models	Recognition of GPS spoofing attacks on UAS
Aissou et al. (2021)	Tree-based supervised ML models	Recognition of GPS spoofing attacks on UAS
Aljabri & Mirza (2022)	ML and deep learning models	Recognition of phishing attacks
Basan et al. (2021)	Kullback–Leibler divergence	GPS spoofing attack Recognition technology
Bose (2022)	Neural network ML	Recognition of GPS spoofing attacks
Dang et al. (2022)	Transfer Learning	GPS spoofing Recognition for Cellular-Connected UAVs
Das et al. (2020)	ML	Novel approach for GPS spoofing Recognition
Jha et al. (2020)	Literature review	Review of methods and technologies for GPS spoofing
Khoei et al. (2022)	Supervised and unsupervised models	Comparative analysis of models for GPS spoofing
Khoei et al. (2022)	Dynamic selection techniques	GPS spoofing attack Recognition
Kwon & Shim (2020)	AHRS/accelerometer	Performance analysis of GPS spoofing Recognition
Lee et al. (2015)	Accelerometers and probability of detection	GPS spoofing Recognition using accelerometers
Pardhasaradhi & Cenkeramaddi (2022)	Distributed radar tracking and fusion	GPS spoofing Recognition and mitigation for drones
Shafiee et al. (2018)	Multi-Layer Neural Network	Recognition of Spoofing Attack using ML
Shafique et al. (2021)	ML models	Recognition of signal spoofing attack in UAVs
Singh et al. (2020)	Literature review	GPS spoofing Recognition and mitigation: A survey
Sung et al. (2022)	1D convolution neural network	GPS spoofing Recognition for small UAVs
Talaei Khoei et al. (2022)	Impact of dataset and model parameters	Performance analysis for GPS spoofing Recognition
Wei et al. (2022)	ML using perception data	UAV GPS spoofing Recognition using perception data
Zuo et al. (2021)	Isolation forest	Recognition of GPS spoofing attacks

Aside from the strategies discussed in the literature review, there are several more intriguing research topics in GPS spoofing detection. Researchers, for example, are looking into ways to leverage several GPS signals to increase the precision and reliability of spoofing detection. They are also working on strategies for detecting spoofing attacks in real time, which is vital for sectors like aviation and marine where fast detection is critical (Jha et al., 2020). As research in this field continues, it is probable that even more efficient strategies for detecting GPS spoofing assaults will be developed. This will aid in the protection of essential infrastructure as well as the safety of persons and property.

EXPERIMENTAL STUDY

In a controlled laboratory environment, GPS signals are examined using SDR and GPS signal generators. The software packages needed for gathering and analyzing GPS signals were then fitted and Connected. For the test, a target GPS receiver is selected. False GPS signals that alter the time and position data are produced using a GPS signal generator or SDR. Simulate several spoofing attacks, including continuous spoofing, spoofing at many locations, and spoofing at a single place. The target receiver's GPS signals were captured and saved using the SDR and monitoring software. Take a close look at the specifics of the signals that were received, such as the timing, signal strength, and signal-to-noise ratio (SNR). Then contrast the signals obtained with what a true GPS signal should look like. Through experimentation, it is possible to analyses GPS signals that have been received, evaluate receiver behavior, use ML algorithms, and validate mitigation strategies. The discoveries will increase our knowledge of GPS spoofing vulnerabilities and aid in the creation of efficient defenses. Signal strength, SNR, and time deviation are three crucial characteristics that we use to determine the effect of GPS spoofing. These properties are thoroughly compared in Table 2 for a variety of situations, including real GPS signals, Single-Site, Multi-Site, and continuous spoofing. The Global Positioning System's authentic transmission is represented by the true GPS signal. It has a timing variation of 0.5 ms, a SNR of 30 dB, and a signal intensity of -130 dBm. These amounts act as the starting point for comparison with the fictitious situations. The simultaneous broadcast of fake GPS signals from several places is known as Multi-Site spoofing, which makes the identification procedure more challenging.

In this case, the SNR drops to 25 dB and the signal intensity drops to -110 dBm. However, with a value of 0.8 ms, the timing divergence is considerably less than with Single-Site spoofing. This disparity might be attributable to the difficulties the attacker had in synchronizing signals generated from various sources. A persistent and continuing spoofing assault intended to trick GPS receivers is referred to as continuous spoofing. Here, the SNR lowers to 22 dB, the signal intensity drops to -105 dBm, and the timing deviation rises to 1.5 ms. Due to the continuous nature of the attack, the spoof signals are exposed for a longer period of time, which has a more noticeable effect on signal quality (Sharma et al., 2023).

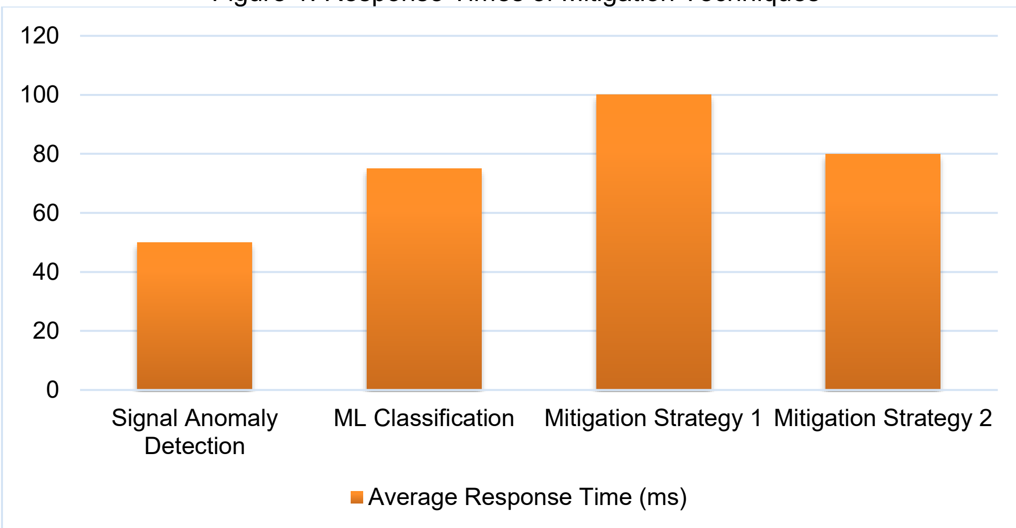
Table 2: Mitigation Strategies Performance Evaluation

Strategies	Detection Accuracy (%)	False Positive Rate (%)	False Negative Rate (%)
Signal Anomaly Detection	92	6	7
ML Classification	97	3	4
Mitigation Strategy 1	-	2	-
Mitigation Strategy 2	-	4	-

RESEARCH OUTCOMES

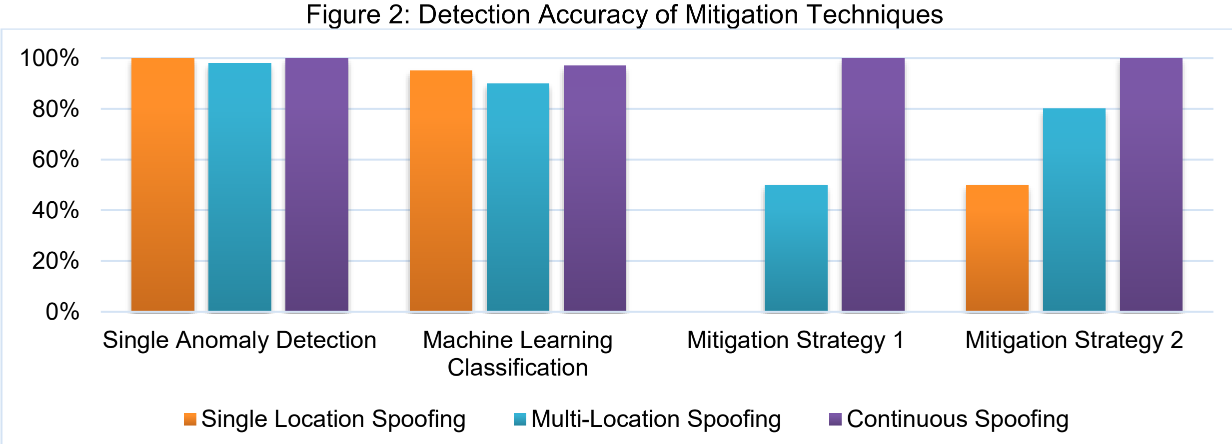
This study evaluates the effectiveness and performance of several methods for identifying and minimizing spoofing assaults as part of its investigation of GPS spoofing detection and mitigation systems. This analysis focuses on the effects and preventative actions used to lessen the risks brought on by GPS spoofing. The mitigating response times are compared in Figure 1. How rapidly each approach can identify and react to spoofing instances is shown by their average response times, which are measured in milliseconds. Response times reveal the effectiveness and quickness of a mitigating strategy. ML classification took 75 ms, whereas signal anomaly detection took 50 ms. Whereas Mitigation Strategy 1 takes100 ms to respond, and Mitigation Strategy 2 did so in just 80 ms.

Figure 1: Response Times of Mitigation Techniques



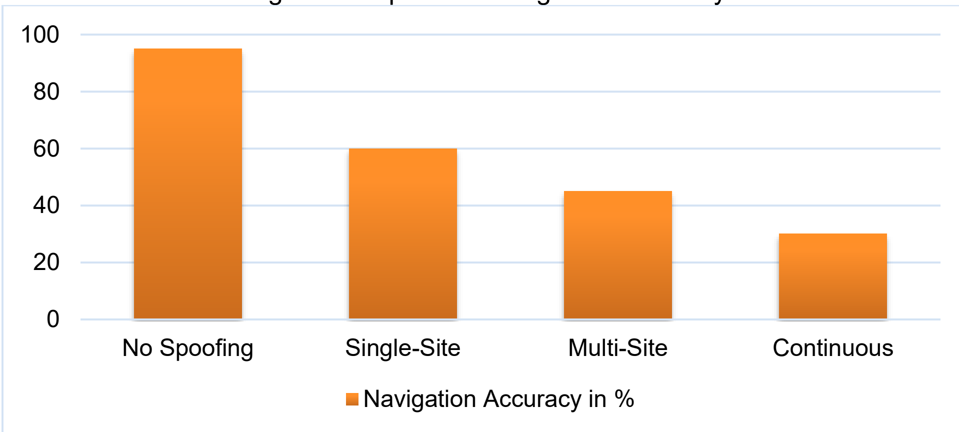
Exploring GPS spoofing detection and mitigation methods aims to give a in-depth knowledge of the efficacy, performance, and reaction times of various solutions. This study contributes significantly to enhancing the security and integrity of GPS-based systems and applications. It serves as a valuable resource for identifying the most reliable and practical methods to detect and counter GPS spoofing attacks effectively.

Figure 2: Detection Accuracy of Mitigation Techniques



Methods for reducing GPS spoofing are evaluated by F2. The ability to recognize Single-Site, multiple-location, and continuous spoofing assaults is evaluated for each approach. All three spoofs were detected by "Signal Anomaly Detection". Using anomaly detection, GPS spoofing signals are found. The adaptability of spoofing detection is shown. In every instance of spoofing, "ML Classification" was successful. GPS spoofing patterns are used to train ML systems, which correctly detect and identify attempts. Due to its high detection rates, ML systems can discriminate between real and fraudulent GPS signals.

Figure 3: Impact on Navigation Accuracy



The presented results underscore the significant adverse impact of GPS spoofing on navigational accuracy, underscoring the critical importance of implementing effective mitigation strategies to mitigate these effects. Our proposed ML-based solution exhibits robust performance with high detection rates across various spoofing scenarios. Specifically, it achieves an impressive 98% detection rate for continuous spoofing, 90% for multi-site spoofing, and 95% for single-site spoofing. These findings highlight the solution's proficiency in discerning and categorizing falsified signals, enabling both precise identification and confident classification of spoofing incidents.

Table 3: Analysing mitigation techniques for several spoofing situations

Techniques	Single-Site Spoofing	Multi-Site Spoofing	Continuous Spoofing
Signal Anomaly Detection	Spotted	Spotted	Spotted
ML Classification	Spotted	Spotted	Spotted
Mitigation Strategy 1	Not efficient	Moderately efficient	Efficient
Mitigation Strategy 2	Moderately efficient	Efficient	Efficient

GPS spoofing mitigation techniques are compared in Table 3, with "Detected" denoting successful detection and "Not efficient" denoting unsuccessful detection. According to the spoofing circumstance, several tactics, such as "Mitigation Strategy 1" and "Mitigation Strategy 2," were successful, according to the research. A was efficient against continuous and multiple-location spoofing, but not spoofing at a single site. When it came to Multi-Site and continuous spoofing, Strategy B was 100% efficient while only being 50% efficient when it came to Single-Site spoofing. The study emphasizes how crucial it is to choose the best mitigation strategy based on the spoofing incident in order to prevent GPS spoofing. The evaluation aids in determining each strategy's efficacy in various spoofing situations.

CONCLUSION AND FUTURE SUGGESTIONS

Overcoming issues caused by GPS spoofing attacks requires knowledge of the detection and mitigation techniques that have been the subject of research. As verification of their capacity to recognize and address GPS spoofing problems, techniques for signal anomaly detection and ML classification performed well in recognizing spoofing attacks in a number of settings. GPS spoofing had a major impact on navigational accuracy, with a noticeable decrease in accuracy when spoofing scenarios were used. A variety of mitigation strategies were more efficient when spoofing continuously at one or more places. Response times were quickest for the "Signal Anomaly Detection" and "ML Classification" approaches. Advancing the field of spoofing detection necessitates further research efforts. These should focus on enhancing the effectiveness of detection algorithms, refining real-time mitigation techniques, exploring innovative multi-sensor fusion strategies, prioritizing user-centric design principles, fortifying GNSS infrastructure security, and establishing comprehensive standardized testing and evaluation frameworks. These critical areas of research will collectively contribute to the ongoing evolution and enhancement of spoofing detection capabilities.

REFERENCES

Aissou, G., Benouadah, S., El Alami, H. E., & Kaabouch, N. (2022). Instance-based supervised machine learning models for detecting GPS spoofing attacks on UAS 12th Annual Computing and Communication Workshop and Conference (CCWC), 2022 (pp. 208–214). IEEE Publications. https://doi.org/10.1109/CCWC54503.2022.9720888

Aissou, G., Slimane, H. O., Benouadah, S., & Kaabouch, N. (2021). Tree-based supervised machine learning models for detecting GPS spoofing attacks on UAS, Electronics y Mobile Communication Conference (UEMCON), 2021 (pp. 649–653). IEEE Publications. https://doi.org/10.1109/UEMCON53757.2021.9666744

Aljabri, M. S., & Mirza, S. (2022). Phishing attacks detection using machine learning and deep learning models 7th International Conference on Data Science and Machine Learning Applications (CDMA), 2022 (pp. 175–180). https://doi.org/10.1109/CDMA54072.2022.00034

Basan, E. S., Basan, A. S., Nekrasov, A., Fidge, C. J., Sushkin, N., & Peskova, O. (2021). GPS-spoofing attack detection technology for UAVs based on Kullback–Leibler divergence. Drones, 6(1). https://doi.org/10.3390/drones6010008

Bose, S. C. (2022). GPS spoofing detection by neural network machine learning. IEEE Aerospace and Electronic Systems Magazine, 37(6), 18–31. https://doi.org/10.1109/MAES.2021.3100844

Dang, Y., Benzaïd, C., Taleb, T., Yang, B., & Shen, Y. (2022). Transfer Learning based GPS Spoofing Detection for Cellular-Connected UAVs, 2022, 629–634. https://doi.org/10.1109/IWCMC55113.2022.9824124

Das, T. K., Pal, S. K., & Sarkar, S. K. (2020). A novel approach for GPS spoofing detection using machine learning. IEEE Transactions on Industrial Informatics, 16(1), 566–575.

Jha, M. K., Singh, A. K., & Jain, V. K. (2020). GPS spoofing: A review of methods and technologies. Sensors, 20(20), 5607.

Khoei, T. T., Gasimova, A., Ahajjam, M. A., Shamaileh, K. A., Devabhaktuni, V. K., & Kaabouch, N. (2022). A comparative analysis of supervised and unsupervised models for detecting GPS spoofing attack on UAVs IEEE International Conference on Electro Information Technology (eIT), 2022 (pp. 279–284). https://doi.org/10.1109/eIT53891.2022.9813826

Khoei, T. T., Ismail, S. S., & Kaabouch, N. (2022). Dynamic selection techniques for detecting GPS spoofing attacks on UAVs. Sensors (Basel, Switzerland), 22.

Kwon, K. C., & Shim, D. S. (2020). Performance analysis of direct GPS spoofing detection method with AHRS/accelerometer. Sensors, 20(4). https://doi.org/10.3390/s20040954

Lee, J., Kwon, K., An, D. S., & Shim, D. (2015). GPS spoofing detection using accelerometers and performance analysis with probability of detection. International Journal of Control, Automation and Systems, 13(4), 951–959. https://doi.org/10.1007/s12555-014-0347-2

Pardhasaradhi, B., & Cenkeramaddi, L. R. (2022). GPS spoofing detection and mitigation for drones using distributed radar tracking and fusion. IEEE Sensors Journal, 22(11), 11122–11134. https://doi.org/10.1109/JSEN.2022.3168940

(A complete list of references is available upon request.)

DECISION SCIENCES INSTITUTE

A Quantitative Examination of the World's Political Systems

Tuncay Bayrak
Western New England University
College of Business
Springfield, MA.
tbayrak@wne.edu

ABSTRACT

While a country's global ranking may be assessed by measures of their GDP, population size, average disposable income, and some other metrics, in this study we examined 166 countries with respect to their regime types and three indices of corruption index, economic freedom index, and happiness index using the data retrieved from several official databases. Our analyses revealed that while there are some similarities among them, the four types of regimes are significantly different from one another in terms of the three indices.

KEYWORDS: Corruption Index, Happiness index, Economic Freedom Index, Political Regimes

INTRODUCTION

A country's global ranking and reputation in the international arena may be evaluated and assessed using a variety of measures. While one would agree that there is no single agreed-upon measure or index that can be used to assess a country's global ranking or reputation, metrics such as a country's corruption score, happiness index, and economic freedom index have been increasingly employed by researchers conducting international surveys to compile a ranking of nations.

As argued by Tisch (2017), a country's reputation has an unquestionable value. In fact, nations today are increasingly concerned with their reputation relative to other nations and turn to actively measuring and managing that reputation (Passow et al., 2005). Consequently, nations and their leaders across the globe strive to overcome a negative image or reputation and build a positive global reputation as a negative global reputation may have implications for foreign policy, tourism marketing, and international trade (Tisch, 2017).

While political systems across the globe may be classified from a variety of perspectives, the various types of political systems found in the sovereign countries of the world are classified as full democracies, flawed democracies, hybrid regimes, and authoritarian regimes (EUI, 2021). One would argue that full democracies or countries with more democratic institutions are perceived as having a strong reputation. Accordingly, the most reputable nations may be perceived as having lower corruption scores, a higher economic freedom index, and a higher happiness index. While these perceptions may help one enhance their understanding of the political systems across the globe, a comparative quantitative analysis of the political systems with respect to their corruptions level, happiness index, and economic freedom score may, therefore, serve as a more credible means of compiling a ranking of nations.

This study aims to provide a quantitative examination of the world's political systems found in 166 nations as classified by the Economist magazine (EUI, 2021). In particular, this study investigates if there are any similarities and statistically significant differences among 166 nations categorized into political systems with respect to their corruption score, happiness index, and economic freedom index.

LITERATURE REVIEW

It may be argued that consciously or unconsciously people are influenced by their perceptions of each country's reputation. These perceptions may be based on their direct experience, stereotypes, and the country's actions. Since a strong reputation may mean more exports, more investments, and more people coming to visit (Valet, 2019), increasing attention has been paid to the relationship between a country's reputation and the factors such as its corruption score that may influence how a country is perceived.

Corruption, defined as the "misuse of public power" (Uroos et al., 2021), is generally known as having a negative impact on a country's reputation and image, and its impact on a country has been studied in the literature. For instance, a study conducted by (Mauro, 1995) holds that corruption is found to lower investment, thereby lowering economic growth. In fact, over the past two centuries, corruption has had a significant negative effect on steady-state growth in various nations (Uberti, 2022). A similar study carried out by (Mo, 2001) maintains that a 1% increase in the corruption level reduces the growth rate by about 0.72%. Mo (2001) further claims that "the most important channel through which corruption affects economic growth is political instability, which accounts for about 53% of the total effect."

As for the factors triggering corruption, Torrez (2002) argues that trade restrictions may cause corruption as they may shift resources from directly productive activities. From a slightly different vantage point, Jong-Sung and Khagram (2005) claim that income inequality increases the level of corruption through material and normative mechanisms. The authors further argue that inequality also adversely affects social norms about corruption and people's beliefs about the legitimacy of rules and institutions, thereby making it easier for them to tolerate corruption as acceptable behavior. Moreover, corruption is found to strongly reduce exports (Charoensukmongkol and Sexton, 2011), and have more long-term negative impacts on developing countries by discouraging bilateral trade (Narayan et al., 2021). While factors such as literacy rate and GDP growth have a negative effect, inflation is found to have a positive effect on corruption (Uroos et al., 2021). The global position of a county in trade can also be affected by the pervasiveness of the country's corruption index (Charoensukmongkol and Sexton, 2011). Overall, as argued by (Uberti, 2022), while grand corruption, as opposed to petty corruption, is more prevalent in autocracies the negative effect of corruption on growth is substantially larger in democracies than in autocracies.

A country's reputation across the globe has also been evaluated through the ranking of national happiness. While different happiness indices have been presented (Sanza et al., 2018), the ranking of national happiness is an important criterion for world leaders to help guide their public policy. To measure a country's happiness index, social scientists often use what's called a "dystopia" indicator, which reflects the country's own perception of doing better or worse than the hypothetical country Dystopia (Carlsen, 2018).

A number of factors seem to play an important role in influencing the national happiness index. For instance, a study by (Galletta, 2016) found that economic conditions are an important factor in assessing happiness. More precisely, studies suggest that government expenditure as a percentage of gross domestic product positively and significantly influences happiness (Perovic and Golem, 2010). Additionally, studies point out that the gross domestic product per capita in terms of purchasing power parity plays an inferior role (Carlsen, 2018). Wang et al. (2019) agree, arguing that an increase in average consumption has a positive effect on happiness. The same observation was made by (Norberg, 2010), claiming that GDP growth correlates with happiness.

In addition to the corruption index and the happiness index, the economic freedom index has been employed to study a country's global reputation and image. The Economic Freedom of the World Index was first produced in 1996 by (Gwartney et al., 1996) and has been updated annually since then. It presented an index that measures the consistency of a nation's policies and institutions with economic freedom (Gwartney and Lawson, 2003). The key ingredients of economic freedom are "personal choice, voluntary exchange, freedom to compete, and protection of person and property." (Gwartney and Lawson, 2003). A similar index was proposed by the Heritage Foundation (HF, 2021) to develop a systematic, empirical measurement of economic freedom in countries throughout the world.

The relationship between the economic freedom index and a country's overall well-being and its reputation has also been studied in the literature. For instance, Carlson and Lundstrom (2002) suggest that economic freedom and the growth of GDP are closely correlated. In general, studies show that economic freedom is a significant determinant of economic growth (Gwartney et al., 1999). In other words, studies suggest a positive correlation between economic freedom and economic growth (Nergren and Jordahl, 2005), better living standards, and more happiness (Hall and Lawson, 2014). Islam (1996) argues that there is a direct relationship between the economic freedom index and per capita income in low-income countries and a direct relationship between the economic freedom index and the growth rate of per capita income in high-income countries. Looking at the issue from a different angle, Heckelman and Knack (2009) investigated the relationship between foreign aid and changes to economic freedom in recipient nations over the 1990–2000 decade. The authors found that foreign aid has no significant effect on economic freedom overall. Pavlic et al., (2022) reached a similar conclusion, arguing that while there is no meaningful causal impact of overall aid on recipient countries' economic freedom, large and sustained increases in "governance-specific" aid lead to modest effects on economic freedom. Finally, Murphy (2022) argues that a broader array of constitutional characteristics and civil liberties have positive effects on economic freedom.

METHODS

This study was conducted using the data published by the Economist Magazine (EIU, 2021), the Heritage Foundation (HF, 2021), Transparency International (TI, 2021), and World Happiness Report (WHR, 2021). The data set utilized in this study contained observations on 166 countries. The overall global reputation ranking of the countries examined in this study was measured using three variables: corruption index, economic freedom index, and happiness index.

Once we retrieved the dataset from the foregoing websites, we processed and analyzed it using SPSS, an advanced multivariate data management and data analytics software application developed by IBM. Throughout this study, a significance level of 0.05 was employed when conducting any statistical test.

In this study, the corruption index is reported on a scale of 0 to 100, where 0 is highly corrupt and 100 is very clean. The economic freedom index is reported on a scale of 0 to 100, where 0 is no economic freedom, and 100 is total economic freedom. Finally, the happiness index is reported on a scale of 0 to 100, where 0 is not happy and 10 is happy.

RESULTS

Descriptive Summaries of The Four Regime Types

The 166 countries examined in this study were first broken down based on their geographic regions. As summarized in Figure 1, while Europe is home to the largest number of countries with full democracies (12), a significant number of authoritarian regimes seem to be located in Africa (27), which is also home to the largest number of hybrid regimes (16), followed by Asia (24), Overall, authoritarian regimes seem to be more prevalent across the globe with 59 (36%) countries classified as authoritarian, followed by 52 (31%) nations categorized as flawed democracies. Out of 166 countries, only 21 (13%) countries were classified as a full democracy.

Figure 1: Distribution of the regimes across geographic regions.

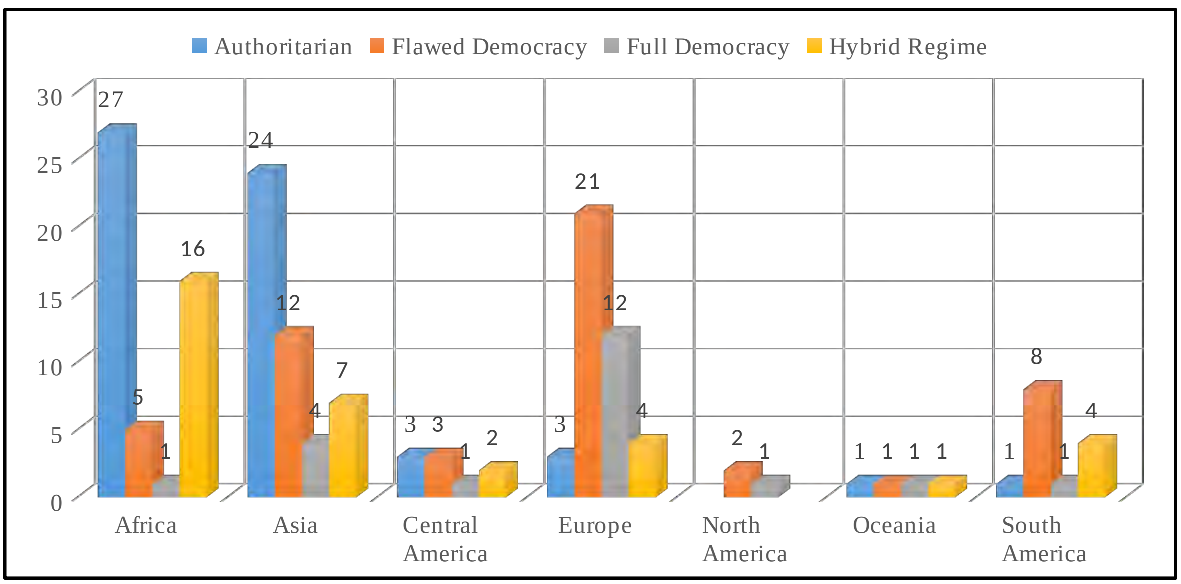
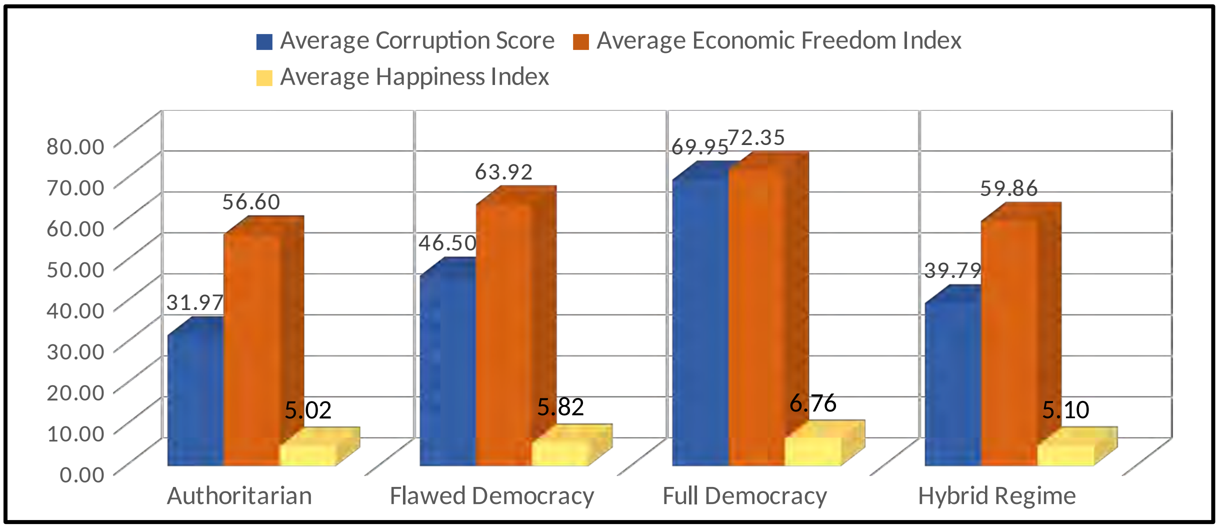


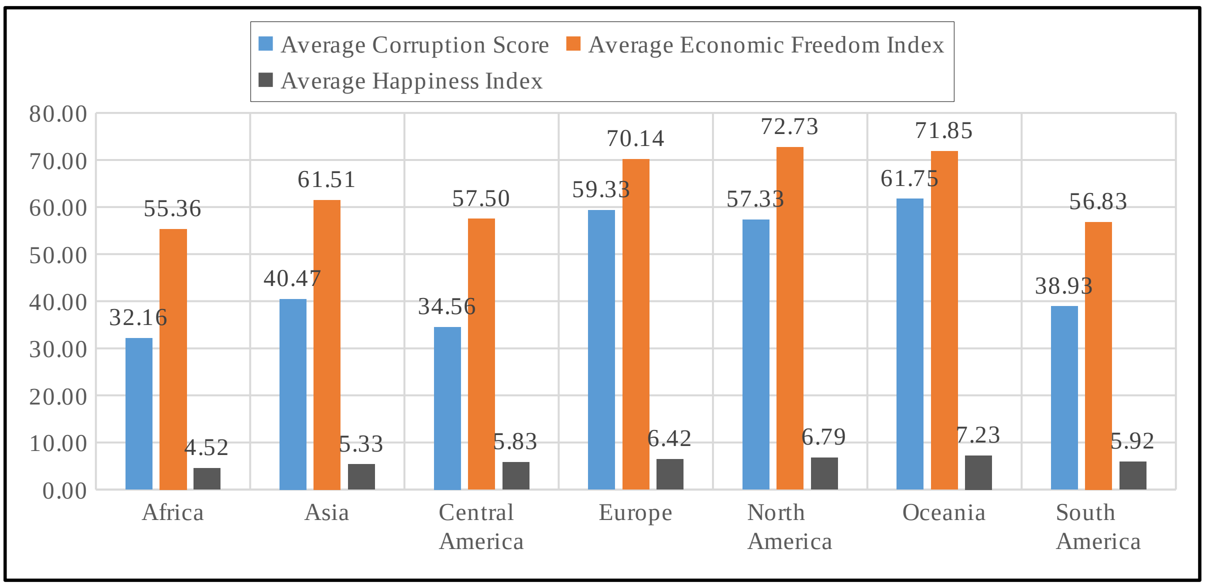
Figure 2 tabulates the average scores of the three indices across the four types of political systems analyzed in this study. As seen, not surprisingly, authoritarian regimes have the worst scores on the three indices followed by hybrid regimes. Across the three indices, full democracy seems to attain the highest average scores. As seen in Table 2, while the authoritarian regimes have the worst average corruption score (31.97), regimes classified as "full democracy" have the best average corruption score (69.95). Similarly, countries with full democracy seem to attain the highest average economic freedom score (72.35), whereas countries with authoritarian regimes offer the lowest average freedom index (56.6). Finally, nations with full democracy seem to be the happiest nations (6.76), followed closely by nations with flawed democracy (5.82). With the lowest average happiness score, countries with authoritarian regimes seem to be the least happy nations.

Figure 2: Regime types and average scores



Scores on three indices across different geographic regions were also analyzed and the results are tabulated in Figure 3, which suggests that while Africa, South America, and Central America have the worst corruption scores, Oceania and Europe have the cleanest records. Similarly, while Oceania and North America have the highest economic freedom index, South America and Africa seem to have the lowest economic freedom index. Finally, while nations in Oceania and North America report the highest happiness index, nations located in Africa report the lowest happiness index.

Figure 3: Regions and average scores



Corruption Score

While numerous factors may lead to vastly different forms of corruption from country to country, to get a better understanding, and to make a comparison among the four regime types, we first conducted an Analysis of Variance (ANOVA) test with respect to their corruption index. Countries around the world are ranked based on perceived levels of public sector corruption. The results are reported on a scale of 0 to 100, where 0 is highly corrupt and 100 is very clean.

The ANOVA test results summarized in Table 1 suggest that since the p-value (0.001) is lower than the level of significance (0.05), the four regime types are significantly different from one another in terms of fighting public sector corruption.

Table 1: Corruption scores and ANOVA test results

ANOVA

Corruption Score

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	23423.689	3	7807.896	35.679	<.001
Within Groups	35451.443	162	218.836		
Total	58875.133	165			

Although the ANOVA test we conducted allows us to determine if there are any statistical differences between the means of the four regime types, it does not show which regime types are significantly different from each other. Therefore, to determine which regimes are significantly different from one another we carried out further post hoc multiple comparisons test using the Tukey method. As seen in Table 2, while in terms of their corruption scores authoritarian regimes and hybrid regimes are similar to one another (p-value=0.071), there is a significant difference between authoritarian regimes and full democracy and hybrid regimes (p-value=0.001). Similarly, there is no significant difference between flawed democracy and hybrid regimes (p-value=0.172). The same output suggests that full democracy is statistically and significantly different from the other types of regimes (p-value=0.001).

Table 2: Multiple comparisons and corruption scores

Multiple Comparisons

Dependent Variable: Corruption Score
Tukey HSD

(I) Democracy Category	(J) Democracy Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Authoritarian	Flawed_Democracy	-14.53390*	2.81380	<.001	-21.8382	-7.2296
	Full_Democracy	-37.98628*	3.75897	<.001	-47.7442	-28.2284
	Hybrid_Regime	-7.82802	3.18519	.071	-16.0964	.4404
Flawed_Democracy	Authoritarian	14.53390*	2.81380	<.001	7.2296	21.8382
	Full_Democracy	-23.45238*	3.82481	<.001	-33.3812	-13.5236
	Hybrid_Regime	6.70588	3.26263	.172	-1.7636	15.1753
Full_Democracy	Authoritarian	37.98628*	3.75897	<.001	28.2284	47.7442
	Flawed_Democracy	23.45238*	3.82481	<.001	13.5236	33.3812
	Hybrid_Regime	30.15826*	4.10574	<.001	19.5002	40.8163
Hybrid_Regime	Authoritarian	7.82802	3.18519	.071	-.4404	16.0964
	Flawed_Democracy	-6.70588	3.26263	.172	-15.1753	1.7636
	Full_Democracy	-30.15826*	4.10574	<.001	-40.8163	-19.5002

*. The mean difference is significant at the 0.05 level.

Tale 3 provides a condensed version of the information presented in the multiple comparisons table (table 2). The homogeneous subsets scores summarized in Table 6 show which pairs of groups have significantly different means on the dependent variable of the corruption index. As seen, authoritarian regimes and hybrid regimes share similarities as they belong to the same subset (subset 1). Similarly, hybrid regime and flawed democracy belong to the same subset (subset 2) as they are similar to one another with respect to their perceived levels of public sector corruption. Finally, full democracy has its own subset (subset 3), suggesting that it's significantly different from the other three types of regimes in terms of their corruption index.

Table 3: Corruption scores homogeneous subsets

Corruption Score

Tukey HSD[a,b]

Democracy Category	N	Subset for alpha = 0.05: 1	2	3
Authoritarian	59	31.9661		
Hybrid_Regime	34	39.7941	39.7941	
Flawed_Democracy	52		46.5000	
Full_Democracy	21			69.9524
Sig.		.121	.230	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 35.332.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Happiness Index

Having compared the four regime types with respect to their corruption scores, an ANOVA test was conducted to determine if there were significant differences among the four regime types with regard to their happiness index.

We postulated that depending on their regime types, nations across the globe would have significantly different happiness levels. The ANOVA test results summarized in Table 4 indeed suggest that there are significant differences among the four political systems with respect to their happiness index (p-value=0.001).

Table 4: Happiness index and ANOVA test results

ANOVA

Happiness Index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.805	3	17.935	22.026	<.001
Within Groups	114.813	141	.814		
Total	168.617	144			

A further analysis was carried out to examine which regime or regimes are significantly different from one another. Table 5 shows that while authoritarian regimes and hybrid regimes are similar to one another with respect to their happiness index (p-value=0.983), both full democracy and flawed democracy have significantly higher happiness indices than authoritarian and hybrid regimes (p-value=0.001).

Table 5: Multiple comparisons and happiness index

Multiple Comparisons

Dependent Variable: Happiness Index
Tukey HSD

(I) Democracy Category	(J) Democracy Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Authoritarian	Flawed_Democracy	-.79412*	.18724	<.001	-1.2809	-.3073
	Full_Democracy	-1.74202*	.23609	<.001	-2.3558	-1.1282
	Hybrid_Regime	-.07688	.20792	.983	-.6175	.4637
Flawed_Democracy	Authoritarian	.79412*	.18724	<.001	.3073	1.2809
	Full_Democracy	-.94790*	.23847	<.001	-1.5679	-.3279
	Hybrid_Regime	.71725*	.21062	.005	.1696	1.2649
Full_Democracy	Authoritarian	1.74202*	.23609	<.001	1.1282	2.3558
	Flawed_Democracy	.94790*	.23847	<.001	.3279	1.5679
	Hybrid_Regime	1.66515*	.25503	<.001	1.0021	2.3282
Hybrid_Regime	Authoritarian	.07688	.20792	.983	-.4637	.6175
	Flawed_Democracy	-.71725*	.21062	.005	-1.2649	-.1696
	Full_Democracy	-1.66515*	.25503	<.001	-2.3282	-1.0021

*. The mean difference is significant at the 0.05 level.

The homogeneous subsets statistics summarized in Table 6 confirm the findings presented in Table 5. Both authoritarian and hybrid regimes belong to the same cluster or subset as they have almost the same happiness level of 5.0. The same table shows that flawed democracy (subset 2) and full democracy (subset 3) have significantly higher average happiness indexes of 5.81, and 6.76, respectively.

Looking at these numbers, it may be argued that countries with highly corrupt regimes generally score lower on the happiness index, while countries with well-established democratic institutions score higher on the happiness index.

Table 6: Happiness index homogeneous subsets

Happiness Index

Tukey HSD[a,b]

Democracy Category	N	Subset for alpha = 0.05 1	2	3
Authoritarian	48	5.0228		
Hybrid_Regime	31	5.0997		
Flawed_Democracy	45		5.8170	
Full_Democracy	21			6.7649
Sig.		.986	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 32.538.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Economic Freedom Index

The Index of Economic Freedom used in this study was developed by the Heritage Foundation (HF, 2021) to provide a systematic and empirical measurement of economic freedom in countries throughout the world.

A similar ANOVA test was performed to investigate if the four regime types differ in terms of their economic freedom index. In general, it may be postulated that countries with full democracy would have a higher economic freedom index, while countries with authoritarian regimes would score lower on the economic freedom index.

As summarized in Table 7, the four types of regimes are significantly different from each other as the p-value (0.001) is less than the level of significance of 0.05.

Table 7: Economic freedom index and ANOVA test results.

ANOVA

Economic Freedom Index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4207.846	3	1402.615	13.031	<.001
Within Groups	17006.877	158	107.638		
Total	21214.723	161			

A further post hoc test reveals that authoritarian regimes and hybrid regimes are similar to one another in terms of their economic freedom index (p-value=0.482). Similarly, flawed democracy and hybrid regimes may be put into the same category as their p-value is not significant (p-value=0.297).

Table 8: Multiple comparisons and economic freedom index.

Multiple Comparisons

Dependent Variable: Economic Freedom Index
Tukey HSD

(I) Democracy Category	(J) Democracy Category	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Authoritarian	Flawed_Democracy	-7.31909*	1.99802	.002	-12.5071	-2.1311
	Full_Democracy	-15.75417*	2.65476	<.001	-22.6475	-8.8609
	Hybrid_Regime	-3.25936	2.27682	.482	-9.1713	2.6526
Flawed_Democracy	Authoritarian	7.31909*	1.99802	.002	2.1311	12.5071
	Full_Democracy	-8.43507*	2.68246	.011	-15.4003	-1.4698
	Hybrid_Regime	4.05973	2.30906	.297	-1.9359	10.0554
Full_Democracy	Authoritarian	15.75417*	2.65476	<.001	8.8609	22.6475
	Flawed_Democracy	8.43507*	2.68246	.011	1.4698	15.4003
	Hybrid_Regime	12.49481*	2.89610	<.001	4.9748	20.0148
Hybrid_Regime	Authoritarian	3.25936	2.27682	.482	-2.6526	9.1713
	Flawed_Democracy	-4.05973	2.30906	.297	-10.0554	1.9359
	Full_Democracy	-12.49481*	2.89610	<.001	-20.0148	-4.9748

*. The mean difference is significant at the 0.05 level.

Table 9 shows that while authoritarian regimes have the lowest economic freedom index (56.59), nations with full democracy have the highest average economic freedom index (72.35). Both authoritarian and hybrid regimes are not significantly different from each other as their average economic freedom indexes are listed under the same subset. Similarly, hybrid regime and flawed democracy are similar to one another as they belong to the same subset (subset 2). Finally, full democracy stands out with the highest average freedom index as it belongs to a totally different category (subset 3).

Table 9: Economic freedom index and homogeneous subsets

Economic Freedom Index

Tukey HSD[a,b]

Democracy Category	N	Subset for alpha = 0.05		
		1	2	3
Authoritarian	56	56.5982		
Hybrid_Regime	33	59.8576	59.8576	
Flawed_Democracy	52		63.9173	
Full_Democracy	21			72.3524
Sig.		.558	.364	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 34.780.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

DISCUSSIONS AND CONCLUSIONS

While people's perceptions of a country may be based on their direct experience, stereotypes, and the country's actions, and events taking place in any country in the world can dramatically influence public opinion about the country, the reputation of a country may be measured and assessed based on a variety of quantitative variables such as corruption index, and happiness index.

It may be argued that most reputable countries usually are perceived as having the lowest corruption levels, better civil liberties, and higher economic freedom, which may generate a tangible economic effect for a country.

While countries may be ranked by measures of their GDP, population size, average disposable income, and some other metrics, in this study we examined 166 countries with respect to their political systems and three indices of corruption index, economic freedom index, and happiness index using the data retrieved from several databases. Our analyses revealed that the four types of regimes are significantly different from one another in terms of the three indices. In regard to their corruption scores, authoritarian and hybrid regimes share similarities as they have the worst corruption scores. With their similar corruption scores, nations with hybrid regimes belong to the same cluster as the nations with flawed democracy. Nations classified as full democracy seem to have the cleanest corruption record. A similar picture arises with respect to the happiness index across the four political systems. Authoritarian and hybrid regimes belong to the same subset with almost the same happiness index. As expected, nations with full democracy reported the highest average level of happiness index of 6.76. As for their economic freedom index, once more, authoritarian and hybrid regimes share similarities. Hybrid regimes also seem to be similar to flawed democracy as they belong to the same cluster. Full democracy appears to have the highest economic freedom index of 72.32.

While authoritarian regimes and hybrid regimes seem to belong to two distinct categories, the line between them is not always clear in terms of their economic freedom index, and corruption scores. A similar picture emerges in comparison between hybrid regimes and flawed democracy.

This study suggests that the best-ranked countries or nations with full democracy receive high scores in all three indices, while the authoritarian regimes and nations with hybrid and flawed democracy tend to receive lower scores in all three dimensions. The indices examined in this study may be used to manage a country’s international reputation to assist public policies designed to improve its reputation in the international arena.

REFERENCES

Berggren, N., and Jordahl, H. (2005). Does free trade really reduce growth? Further testing using the economic freedom index. *Public Choice*, 122, 99–114.

Carlsen, L. (2018). Happiness as a sustainability factor. The world happiness index: a posetic-based data analysis. *Sustainability Science*, 13, 549–571.

Carlsson, F., and Lundström, S. (2002). Economic freedom and growth: decomposing the effects. *Public Choice*, 112(3/4), 335–344.

Charoensukmongkol, P. and Sexton, S. (2011). The Effect of corruption on exports and imports in Latin America and the Caribbean. *Latin American Business Review*, 12(2), 83-98.

EIU (Economist Intelligence Unit). (2021). Democracy Index. https://pages.eiu.com/rs/753-RIQ-438/images/eiu-democracy-index-2021.pdf

Galletta, S. (2016). On the determinants of happiness: a classification and regression tree (CART) approach. *Applied Economics Letters*, 23(2). 121-125.

Gwartney, J.D., Lawson, R.A., and Block, W. (1996). Economic freedom of the world: 1975-1995. www.fraserinstitute.org/sites/default/files/EconomicFreedomoftheWorld1975-1995.pdf

Gwartney, J.D., Lawson, R.A., and Holcombe, R. G. (1999). Economic freedom and the environment for economic growth, *Journal of Institutional and Theoretical Economics*, 155(4). 643-663.

Gwartney, J.D., and Lawson, R. A. (2003). Economic freedom index, The concept and measurement of economic freedom, *European Journal of Political Economy*, 19(3), 405-430.

HF (Heritage Foundation). 2021. Highlights of the 2021 Index of economic freedom. https://www.heritage.org/index/pdf/2021/book/2021_IndexofEconomicFreedom_Highlights.pdf

Hall, J. C., and Lawson, R. A. (2014). Economic freedom of the world: an accounting of the literature, *Contemporary Economic Policy*, 32(1), 1-19.

Heckelman, J. C., and Knack, S. (2009). Aid, economic freedom, and growth, *Contemporary Economic Policy*, 27(1), 46-53.

Islam, S. (1996). Economic freedom, per capita income and economic growth, *Applied Economics Letters*, 3(9), 595-597.

Jong-Sung, Y. and Khagram, S. (2005). A comparative study of inequality and corruption. *American Sociological Review*, 70(1), 136-157.

Mauro, P. (1995). Corruption and growth, *The Quarterly Journal of Economics*, 110(3), 681–712.

Mo, P.H. (2001). Corruption and economic growth, *Journal of Comparative Economics*, 29(1), 66-79

Murphy, R. H. (2022). The constitution of ambiguity: The effects of constitutions on economic freedom. *Economic Affairs*, 42(2), 240-258.

Narayan, S. and Minh, N., and Bui, T. (2021). Does corruption in exporter and importer countries influence international trade?, *Emerging Markets Finance and Trade*, 57(11), 3202-3221

Norberg, J. (2010). GDP and its enemies: The questionable search for a happiness index, *European View*, 9(2), 281-282.

Passow, T., Fehlmann, R, and Grahlow, G. (2005). Country reputation-from measurement to management: The case of Liechtenstein. *Corporate Reputation Review*, 7(4), 309-326.

Pavlik, J.B., Powell, B., and Young, A. T. (2022). Does aid cause changes in economic freedom? *International Studies of Economics,* 89(1), 90-111.

Perovic, L.M., and Golem, S. (2010). Investigating macroeconomic determinants of happiness in transition countries, *Eastern European Economics*, 48(4), 59-75.

Sanza, M.T., Caselles, A., Micóc, J.C., and Soler, D. (2018). A stochastic dynamical social model inv**olvin**g a human happiness index. *Journal of Computational and Applied Mathematics*, 340, 231-246.

Tisch, D. (2017). What Gives a nation a great Reputation? https://policyoptions.irpp.org/magazines/july-2017/what-gives-a-nation-a-great-reputation/#:~:text=Reputation%20has%20an%20unquestionable%20economic, stereotypes %2C%20influences%20and%20direct%20experiences.

Torrez, J. (2002). The effect of openness on corruption. *The Journal of International Trade & Economic Development*, 11(4), 387-403.

Transparency International (TU). 2021. Corruption perception index, https://www.transparency.org/en/cpi/2021

Uberti, L.J. (2022). Corruption and growth: Historical evidence, 1790–2010, *Journal of Comparative Economics*, 50(2), 321-349.
Uroos, A., Shabbir, M.S., Zahid, M.U., Yahya, G., and Abbasi, B.A. (2022) Economic analysis of corruption: evidence from Pakistan, *Transnational Corporations Review*, 14(1), 46-61.

Valet, V. (2019). The world’s most reputable countries 2019, https://www.forbes.com/sites/vickyvalet/2019/10/15/ the-worlds-most-reputable-countries-2019/?sh=1e8193b84cb8.

Wang, H., Cheng, Z., and Smyth, R. (2019). Consumption and happiness, *The Journal of Development Studies*, 55(1), 120-136.

WHR (World Happiness Report). 2021.https://happiness-report.s3.amazonaws.com/ 2021/WHR+21.pdf

DECISION SCIENCES INSTITUTE

A Simheuristic Approach for Repair Kit Inventory Policy Creation

John Maleyeff
Boston University Metropolitan College
Email: maleyeff@bu.edu

Ruthairut Wootisarn
Boston University Metropolitan College
Email: ruthwts@bu.edu

Jia Fang
Boston University Metropolitan College
Email: fangj@bu.edu

Yugesh Asokan
Boston University Metropolitan College
Email: yugesh@bu.edu

ABSTRACT

A simheuristic approach is described that determines a periodic review inventory policy for a repair kit, where the repair cannot begin until all parts necessary for the repair are available. The problem was motivated by a public transportation entity that stores over 3 million repair parts worth $50 million. Inventory optimization needs to account for dependencies across the parts in the repair kit, which causes competitional challenges. The heuristic model calculates slopes of a cost equation starting with a deterministic model for each part. The iterative steps to determine an effective inventory policy are described, and an example is presented.

KEYWORDS: Simheuristics, Repair inventory, Repair kit, Public transportation

INTRODUCTION

Many organizations maintain an inventory of parts used to maintain equipment and other resources that are used by their customers or within their own operations. These parts are referred to as repair parts or spare parts. Their inventory levels can be difficult to plan, especially when demand is random and intermittent. A specific class of repair part inventory problems are referred to as repair kit problems. Here, a set of parts is used to perform a repair and the repair can only take place when all necessary parts are available. The dependencies among the parts in a repair kit make these problems a computational challenge.

This article details a simheuristic approach for determining an effective inventory policy for a repair kit. It was motivated by a recent problem that took place at a public transit system that stores over 20000 different part types that together constitute a total repair part inventory of over 3 million parts and accounts for over $50 million in total inventory value. The simheuristic approach combines an individual part optimization formulation with a repair kit simulation. A deterministic optimization formulation provides the initial feasible solution to the simulation, then a stochastic model acts as the heuristic to calculate incremental improvements based on unit changes to each inventory planning parameter. Its main implementation challenge concerns the random nature of simulation results because in some cases two potential solutions cannot be precisely compared (i.e., a new simulated solution cannot always be shown as being better or worse than a previous best solution).

The article starts by providing a brief literature review of repair kit inventory optimization and the simheuristic approach that focuses on its use for operational problem solving. The methodology section describes the repair kit simulation and derives the heuristic model used to search for improved inventory policies. The simheuristic application is then described in detail, then two examples show its application. The simheuristic algorithm requires a better stopping rule, which constitutes future work on this project.

LITERATURE REVIEW

Research on repair inventory systems has been conducted across various industries and disciplines. Kapoor and Ambekar (2015) and Mabini (2016) have conducted literature reviews on repair inventory systems, while Boddupalli et al (2019) and Sun et al (2020) have focused on repair part inventory management and safety stock policies, respectively, in public transit systems. Lin et al (2017) have explored mathematical programming models for repair inventory systems in rapid-speed trains, while Corazza et al (2018) have examined monitoring approaches for repair inventory systems in buses. Repair systems are important in many public settings because these entities maintain equipment for many years. Kim et al (2007), Basten and Houtum (2014) and Slay and Sherbrooke (1988) conducted research focused on public settings, with a particular emphasis on the Coast Guard, Air Force, and radar systems. Diaz and Fu (1997) also explored repair inventory management specifically in public transportation systems.

The repair kit problem occurs when the parts needed for a repair are collected and the repair takes place only after all of the parts needed for the repair are available (Brumelle and Granot, 1993). Because repairs often have intermittent demand, managing repair kit inventory has been done using a one-for-one replenishment policy (Mamer and Smith, 1985). Optimizing repair kit inventory policies is a complex undertaking when cost minimization is sought (Hu et al, 2018). Teunter (2006) developed two heuristics models, one that set a service level and minimized cost and the other that assumed a fixed penalty cost when a repair could not be fulfilled. Prak et al (2017) also presented a heuristic approach that calculated an economic order quantity to represent the difference between reorder points and order-up-to levels in a periodic review system.

Research on the forecasting demand for repair parts usually emphasizes the intermittent nature of repair demand. Kennedy et al (2002) and Kapoor and Ambekar (2015) have studied repair part demand based on maintenance schedules or the occurrence of random failures. Demand forecasting has been examined using different models such as the Poisson model by Silver et al (2016), the exponential model by Hollier (1980), and the mixed zero-truncated Poisson model by Jiang et al (2020). Slay and Sherbrooke (1988) have also discussed the limitations of the Poisson assumption in practice. Schuh et al (2015) have used a Weibull model to predict repair demand based on failure probabilities.

Simheuristics is an approach that blends simulation and heuristics to solve complex stochastic combinatorial optimization problems. The simulation component of simheuristics is used to model stochastic processes and the heuristics component is used to discover good solutions in the solution space. Juan et al (2015) provide an overview of the essential aspects of simheuristics and its applications to a diverse set of scenarios. These authors emphasize the benefits of utilizing simheuristics, including its focus on robustness and adaptability. Simheuristics has been shown to be effective in addressing a wide range of optimization challenges, but finding optimal solutions in reasonable computing time remains elusive (Rabe et al, 2020).

Simheuristics is used for applications in manufacturing and operations, including inventory control, transportation planning, and scheduling (Michalak and Knowles, 2016). Arts et al (2016) applied simheuristics in a changing demand environment to determine repairable stocking and expediting strategies for an annealing manufacturing process. Seiringer et al (2022) provide an example of simheuristics applied to the optimization of a materials requirements planning system. They use a simulation model to represent the production system and a genetic algorithm to search for an optimal solution. Going forward, Juan et al (2015) suggest that new research focus on improving the performance and scalability of simheuristics algorithms, as well as exploring new applications and problem domains.

METHODOLOGY

The simheuristic algorithm applied here attempts to identify a good set of period review inventory policies for each part in a repair kit. The parameters of interest are the reorder point and the order-up-to level for each part that are not consistent across parts due to their costs and lead times. The approach includes a repair kit simulation, a heuristic model, and a search procedure. The simulation model is a robust formulation that mimics a repair kit, and includes fixed ordering costs, inventory holding costs, and costs for delaying a repair. The initial simulation run uses a simplified part-by-part deterministic model that accounts for delays by quantifying the cost of holding other parts in a repair kit when a given part is unavailable. The heuristic model calculates the incremental cost when each of a part's parameters (i.e., its reorder point and order-up-to level) are increased or decreased by 1. After identifying the parameter change that results in the largest negative cost difference, the modification is made to the inventory policy and the simulation is executed. This procedure continues as better solutions are generated and terminates when a poor solution is found. Each of these components is described in more detail below.

Simulation Model

The logic underlying the repair kit simulation is shown as Figure 1. It mimics a repair kit over a specified number of iterations (i.e., weeks). For each trial of the simulation, weekly repair demand is generated from a Poisson distribution. The demand plus backlog from previous weeks is compared to the minimum of all part inventories (i.e., the number of complete repair kits available). Based on this comparison, the number of repair kits that can be shipped to garages is determined, and the delays (in weeks) for each repair kit shipped is recorded.

```
Simulation Setup:
    Input repair kit parameters: Demand/week
    Input parts parameters: Lead time, Part cost, Order-up-to level, Reorder point
    Input cost parameters: Delay cost/week, Holding rate/week, Setup cost/order
    Input previous best solution: Sample mean, Sample standard deviation, Sample size
    Input simulation timeframe: Trials, Iterations (i.e., Weeks), Warmup iterations
    FOR each Trial DO:
        Initialize part information: Inventory, Receive, Pipeline, Reserve, Shipped
        Initialize delay information: Weeks Delay, Queue arrivals, Queue departures
        Initialize cost information: Holding rate/week, Setup cost/order, Delay cost/week
        cycle_service_level = fill_rate = reserve = received = 0
        FOR each Week DO:
            Generate repair_demand (Poisson)
            Calculate available = minimum of all part inventories
            Calculate needed = repair_demand + reserve
            Calculate shipped = minimum of available & needed
            Update arrival and departure lists for shipped parts
            FOR each Part DO:
                Calculate end_inv = begin_inv - shipped
                Update pipeline = pipeline – received
                Order = order_up_to – end_inventory - pipeline
                If end_inv + pipeline >= reorder_point then order = 0
                Update pipeline = pipeline + order
                Update receive = order (lead_time weeks ago)
                Update holding_cost/week by end_inv * holding_rate * part_cost
                if order > 0 then update setup_cost/week by setup_cost/order
                Update begin_inv = end_inv + received
            END FOR
            Update delay_cost/week for shipped repair kits
            Calculate cycle_service_level and fill_rate
            Calculate reserve = reserve + demand - shipped
        END FOR
        Annualize holding_cost, setup_cost, delay_cost
        Calculate total_cost
        Accumulate cost and performance statistics
    END FOR
    Calculate confidence interval for holding_cost, setup_cost, delay_cost
    Perform hypothesis test to compare with previous best solution
    Calculate slopes for increasing/decreasing reorder_point and order_up_to-level for each part
    Sort slopes in ascending order
```

Figure 1: Repair Kit Simulation

The simulation then considers each part in the repair kit. Their inventory levels are updated based on the number of repair kits shipped, then each part's inventory level is compared to the reorder point. If a part's inventory is less than the reorder point, an order is placed for the difference between the order-up-to level and the inventory and the fixed ordering (i.e., setup) cost is recorded. The week number for which the order will be received is stored. Finally, each part's inventory is updated with the parts received from suppliers and inventory holding cost is updated.

After the number of specified iterations, the warmup period data are removed, and the delay costs are calculated. Once all trials are completed, 95% confidence intervals are calculated for the setup, holding, delay, and total costs. The simheuristic model is then applied to calculate the "slopes" associated with making unit changes of one unit to each part's reorder point and order-up-to-level. These values are printed so that the user can manually change the inputs for the subsequent runs of the simulation, until a stopping condition is encountered.

Heuristic Modeling

The heuristic model calculates the estimated improvements (i.e., total cost reductions) for each potential change as follows (where s is the part's reorder point and S is the part's order-up-to level):

1. Decreasing the part's reorder point by 1 ($s - 1$)
2. Increasing the part's reorder point by 1 ($s + 1$)
3. Decreasing the part's order-up-to level by 1 ($S - 1$)
4. Increasing the part's order-up-to level by 1 ($S + 1$)

The heuristic model assumes that demand is Poisson, the lead time is constant, the order cycle is fixed, and the beginning cycle inventory is fixed. The inventory policy is periodic review (s, S) with backlogged demand. Orders are placed weekly. The notation used to implement the model is shown below:

μ = Mean demand per week
L = Deterministic lead time in weeks
I = Inventory after part receipt
T = Cycle duration (weeks)
C_S = Fixed ordering (i.e., setup) cost
C_D = Delay cost per week
C_P = Part purchase cost
h = Annual carrying cost rate
50 = Number of weeks per year

To be feasible, we assume that the order up-to-level is sufficient (i.e., $S \geq \mu L$). Figure 2 shows the fixed order cycle, where x is equal to the demand during the order cycle. The expected cycle length (T) and expected starting inventory (I) per cycle are shown in Equations 1 and 2.

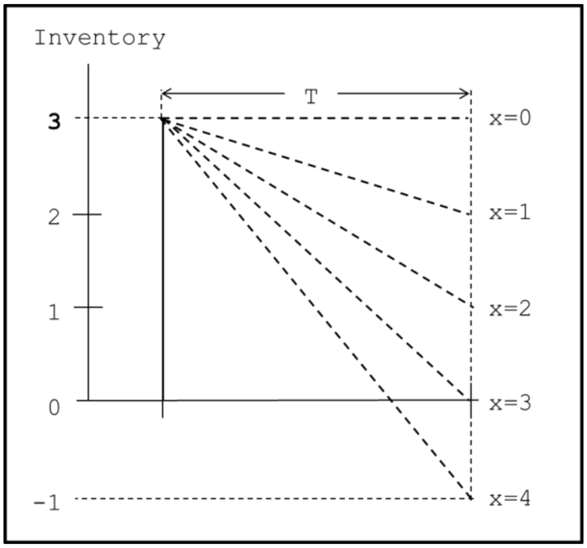


Figure 2: Example Ordering Cycle

$$T = \frac{S - s + 1}{\mu} \tag{1}$$

$$I = S - \mu L \tag{2}$$

The incremental estimated reductions (i.e., the “slopes”) for the total annual cost are calculated by determining the changes in the annual delay, holding, and setup costs for each of the four potential (s, S) inventory policy changes. These calculations are done for each part in the repair kit.

Delay Cost Slope Determination

Delay cost slopes are determined using the Poisson distribution (with mean λ) to calculate probabilities associated with demand during the fixed cycle time. Equation 3 shows the expected annual delay cost for the current best solution. The basis for this derivation is that a delay occurs when cycle demand exceeds I. The derivation assumes that, in these cases, the demand is evenly distributed within the cycle.

$$DC_0 = 50C_D \sum_{d=I+1}^{\infty} \frac{P(X = d|\lambda = \mu T)}{2d} \sum_{t=1}^{d-I} (2t - 1) \tag{3}$$

Equations 4 to 7 show the expected annual delay cost when the four possible changes are made for a part. The differences between each of these costs and DC_0 constitutes the slope of the change. For example, $DC_{S+1} - DC_0$ is the estimated slope in the annual delay cost when S is increased by 1.

$$DC_{S+1} = 50C_D \sum_{d=I+2}^{\infty} \frac{P(X = d|\lambda = \mu T + 1)}{2d} \sum_{t=1}^{d-I-1} (2t - 1) \tag{4}$$

$$DC_{S-1} = 50C_D \sum_{d=I}^{\infty} \frac{P(X = d|\lambda = \mu T - 1)}{2d} \sum_{t=1}^{d-I+1} (2t - 1) \tag{5}$$

$$DC_{s+1} = 50C_D \sum_{d=I+1}^{\infty} \frac{P(X = d|\lambda = \mu T - 1)}{2d} \sum_{t=1}^{d-I} (2t - 1) \tag{6}$$

$$DC_{s-1} = 50C_D \sum_{d=I+1}^{\infty} \frac{P(X = d|\lambda = \mu T + 1)}{2d} \sum_{t=1}^{d-I} (2t - 1) \tag{7}$$

Holding Cost Slope Determination

Equation 8 shows the expected annual holding cost for the current best solution. The basis for this derivation is that the inventory level may be positive for the entire cycle (when $I \geq \mu T$) or inventory may be positive for only a portion of the cycle ($I < \mu T$).

$$HC_0 = \begin{cases} \dfrac{hC_P(2I - \mu T)}{2}, & I \geq \mu T \\ \dfrac{hC_P I^2}{2\mu T}, & I < \mu T \end{cases} \tag{8}$$

Equations 9 to 12 show the expected annual holding cost when the four possible changes are made for a part. The differences between each of these costs and HC_0 constitutes the slope of the change. For example, $HC_{S+1} - HC_0$ is the estimated unit increase (slope) in the annual holding cost when S is increased by 1.

$$HC_{S+1} = \begin{cases} \dfrac{hC_P[2(I + 1) - (\mu T + 1)]}{2}, & I \geq \mu T \\ \dfrac{hC_P(I + 1)^2}{2(\mu T + 1)}, & I < \mu T \end{cases} \tag{9}$$

$$HC_{S-1} = \begin{cases} \dfrac{hC_P[2(I - 1) - (\mu T - 1)]}{2}, & I \geq \mu T \\ \dfrac{hC_P(I - 1)^2}{2(\mu T - 1)}, & I < \mu T \end{cases} \tag{10}$$

$$HC_{s+1} = \begin{cases} \dfrac{hC_P[2I-(\mu T-1)]}{2}, & I \geq \mu T \\ \dfrac{hC_P I^2}{2(\mu T-1)}, & I < \mu T \end{cases} \tag{11}$$

$$HC_{s-1} = \begin{cases} \dfrac{hC_P[2I-(\mu T+1)]}{2}, & I \geq \mu T \\ \dfrac{hC_P I^2}{2(\mu T+1)}, & I < \mu T \end{cases} \tag{12}$$

Setup Cost Slope Determination

Equation 13 shows the expected annual setup cost for the current best solution.

$$SC_0 = \frac{50C_S}{T} \tag{13}$$

Equations 14 to 17 show the expected annual setup cost when the four possible changes are made for a part. The differences between each of these costs and SC_0 constitutes the slope of the change. For example, $SC_{S+1} - SC_0$ is the estimated unit increase (slope) in the annual setup cost when S is increased by 1.

$$SC_{S+1} = \frac{50\mu C_S}{\mu T+1} \tag{14}$$

$$SC_{S-1} = \frac{50\mu C_S}{\mu T-1} \tag{15}$$

$$SC_{s+1} = \frac{50\mu C_S}{\mu T-1} \tag{16}$$

$$SC_{s-1} = \frac{50\mu C_S}{\mu T+1} \tag{17}$$

The total cost difference is the sum of the delay, holding, and setup cost slopes for each potential change.

Search Procedure

The search procedure is described in Figure 3. It is initiated using a deterministic model that generates a (s, S) inventory policy for each part in a repair kit (Maleyeff et al, 2022). These reorder points and order-up-to levels are input to the simulation, which generates results for Run 0 of the simulation and labels this result the best solution. The slopes (described above) are calculated by the simulation program and sorted. Next, the parameter change associated with the most negative total cost slope is made, and the revised part parameters are input to the simulation (i.e., Run 1). Simulation results are compared to the best solution using a hypothesis test that accounts for simulation output variation, and again slopes are calculated.

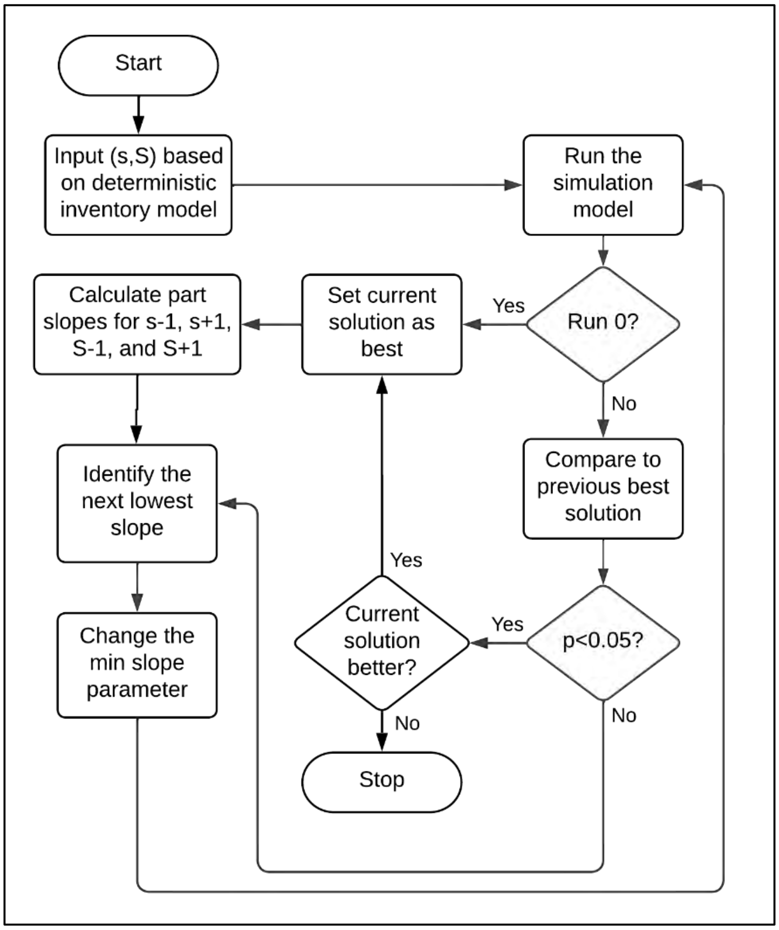


Figure 3: Search Procedure

If the hypothesis test comparing the current solution to the best solution concludes with confidence that the latest run of the simulation resulted in a better solution, its parameters are set as the best solution. Slopes are calculated for changes in each parameter and the procedure continues by simulating the new set of parameters. If the hypothesis test is inconclusive, then the best solution is not updated, but the slopes are calculated and sorted, and the parameters are changed for the next simulation run. If the hypothesis test concludes with confidence that the current solution is worse than the best solution, the procedure terminates. This stooping rule is preliminary as ideas are sought for a more robust stopping rule.

RESULTS

The simheuristics procedure is illustrated using two examples. The repair kit consist of 4 parts (A, B, C, and D), with costs of $10, $50, $100, and $240 per part. The example assumes that the repair delay cost is $5 per week, holding rate is 24% per year, the setup cost is $20 per order, and the weekly demand is 0.4 repairs per week. Each part's lead time is assumed to be 10 weeks in the first example and 20 weeks in the second example. For every simulation run, there were 400 trials, 5000 weeks per trial, and 100 weeks for the warmup period. The number of trials was chosen so that the standard error of the total annual cost was less than 1% of its average. The warmup period was confirmed using Welch's method (Welch, 1983, pp. 268-328).

Example 1

The deterministic model generated values for the periodic review inventory policy, as shown in the Run 0 column in Figure 4. The total annual cost for this policy had a 95% confidence interval of between \$1392 and \$1423. The minimum slope for this scenario was -\$74.45, corresponding to an increase in the reorder point for part D. The simulation of this policy (Run 1 column in Figure 4) had a 95% confidence interval of between \$1054 and \$1078. This was an improvement in total cost ($p<0.05$). The minimum slope for this scenario was -\$56.86, corresponding to another increase in the reorder point for part D. The simheuristic approach continued as shown in Figure 4. The procedure stopped after Run 8, and the best solution is shown as the Result column. The 95% confidence interval for the total annual cost of this policy was between \$776 and \$782.

Part		Run order									Result
		0	1	2	3	4	5	6	7	8	
A	s	5				+	+		+	+	7
	S	23									23
B	s	5						+			6
	S	12									12
C	s	5									5
	S	10									10
D	s	1	+	+	+						4
	S	7									7
Average cost		1407	1066	843	829	825	820	779	783	784	779

Figure 4: Simheuristic Approach (Example 1)

Figure 5 shows the breakdown of the estimated annual holding, setup, and delay costs. Figure 6 shows the 95% interval for the total cost (only Runs 2 through 8 are shown).

Example 2

With increased lead times (from 10 to 20 weeks for each part), the deterministic model generated values for the periodic review model, as shown in the Run 0 column in Figure 7. The total annual cost for this policy had a 95% confidence interval of between \$921 and \$961. The minimum slope for this scenario was -\$13.46, corresponding to an increase in the reorder point for part A. The simulation of this policy (Run 1 column in Figure 7) had a 95% confidence interval of between \$888 and \$912. This was an improvement in total cost ($p<0.05$). The minimum slope for this scenario was -\$9.19, corresponding to an increase in the reorder point for part A. The simheuristic approach continued as shown is Figure 7. The procedure stopped after Run 6, and the best solution is shown as the Result column. The 95% confidence interval for the total annual cost of this policy was between \$860 and \$874.

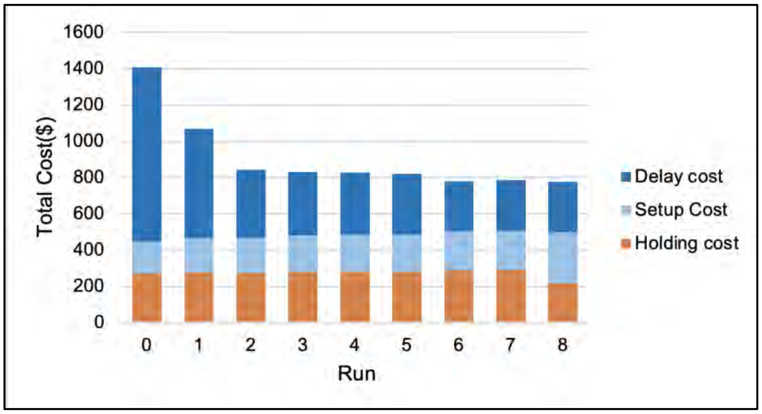


Figure 5: Cost Comparisons (Example 1)

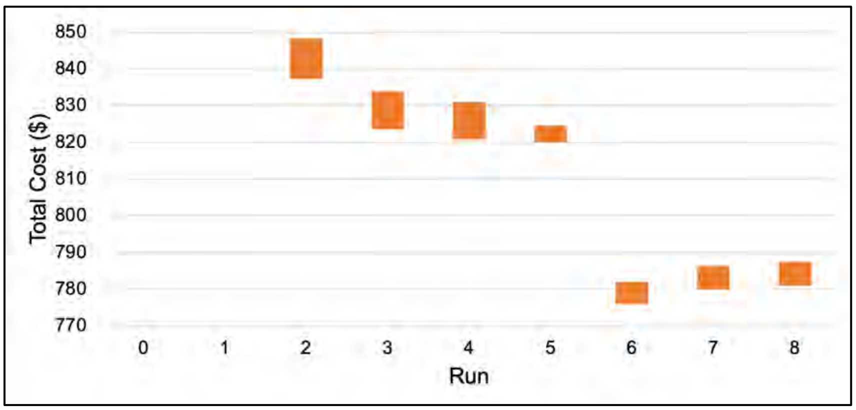


Figure 6: Total Cost Confident Intervals (Example 1)

Part		Run order							Result
		0	1	2	3	4	5	6	
A	s	9	+	+		+	+		11
	S	27							27
B	s	9			+				10
	S	16							16
C	s	9							9
	S	14							14
D	s	9						+	9
	S	12							12
Average cost		941	900	896	867	864	866	908	867

Figure 7: Simheuristic Approach (Example 2)

Figure 8 shows the breakdown of the estimated annual holding, setup, and delay cost.

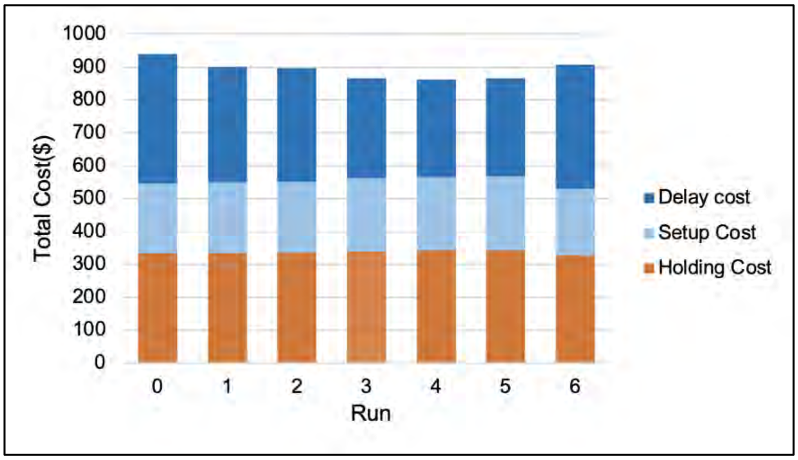


Figure 8: Cost Comparisons (Example 2)

Figure 9 shows the 95% interval for the total cost.

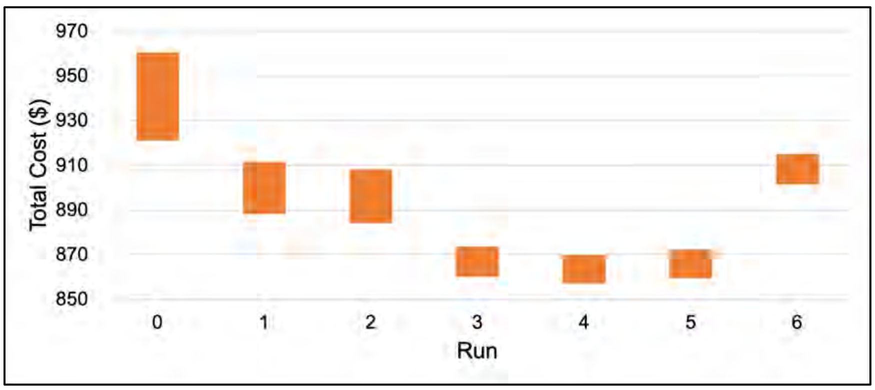


Figure 9: Total Cost Confident Intervals (Example 2)

FUTURE WORK

The simheuristic approach is appropriate for determining inventory policies for the repair kit problem because it is computationally challenging, and heuristics can be developed that generate improved solutions. The approach presented in this article appears effective at improving inventory policies in the period review case. The authors have more work to do for two important aspects of the simheuristics algorithm. First, the search procedure needs to be more robust by not stopping at the first poor solution. A modified branch-and-bound procedure is being considered. Second, the simheuristic algorithm needs to be validated. Given the computational burdens, it would be impossible to validate the algorithm by comparing with an exact solution and therefore this challenge remains elusive.

ACKOWLEDGEMENTS

This work, which commenced in September 2021 as a collaboration with the MBTA, was undertaken by the Boston University Metropolitan College Decision Sciences Laboratory. The authors wish to acknowledge the work of other researchers who made significant contributions to the development of the concepts and models that form the foundation of the simheuristic approach - Ningyi Yang, Jiehua Gong, Jingran Xu, Alex Yang, Valeria Kaffaty, and Bob Jiang.

REFERENCES

Arts, J., Basten, R., & Van Houtum, G. (2016). Repairable stocking and expediting in a fluctuating demand environment: Optimal policy and heuristics. *Operations Research, 64*(6), 1285-1301.

Basten, R.J.I. & Van Houtum, G.J. (2014). System-oriented inventory models for spare parts. *Surveys in Operations Research and Management Science, 19*, 34-55.

Boddupalli, S.S., Sherman, A., Zerkus, J., & Grossman, A. (2019). Maintrainenance: Keeping heavy rail maintenance on track. *Transportation Research Record, 2673*(7), 192-203.

Brumelle, S. & Granot, D. (1993). The repair kit problem revisited. *Operations Research, 41*(5), 994-1006.

Corazza, M.V., Magnalardo, S., Musso, A., Petracci, E., Tozzi, M., Vasari, D., & de Verdalle, E. (2018). Testing an innovative predictive management system for bus fleets: outcomes from the Ravenna case study. *IET Intelligent Transport Systems, 12*(4), 286-293.

Díaz, A. and Fu, M.C. (1997). Models for multi-echelon repairable item inventory systems with limited repair capacity, *European Journal of Operational Research, 97*, 480-492.

Hollier, R.H. (1980). The distribution of spare parts. *International Journal of Production Research, 18*(6), 665-675.

Hu, Q., Boylan, J.E., Chen, H., & Labib, A. (2018). OR in spare parts management: A review. *European Journal of Operational Research, 266*(2), 395-414.

Jiang, A., Tam, K.L., Guo, X., & Zhang, Y. (2020). A new approach to forecasting intermittent demand based on the mixed zero-truncated Poisson model. *Journal of Forecasting, 39(1)*, 69-83

Juan, A.A., Faulin, J., Grasman, S.E., Rabe, M., and Figueira, G. (2015). A review of simheuristics: Extending metaheuristics to deal with stochastic combinatorial optimization problems. *Operations Research Perspectives, 2*, 62-72.

Kapoor, R., & Ambekar, S. (2015). Closed loop repairable parts inventory system: A literature review. *Indian Journal of Economics & Business 14*(1), 31-47.

Kennedy, W.J., Wayne Patterson, J., & Fredendall, L.D. (2002). An overview of recent literature on spare parts inventories. *International Journal of Production Economics, 76*(2), 201-215.

Kim, J., Kim, T., & Hur, S. (2007). An algorithm for repairable item inventory system with depot spares and general repair time distribution. *Applied Mathematical Modelling, 31*(5), 795-804.

Lin, B., Wang, J., Wang, H., Wang, Z., Li, J., Lin, R., Xiao, J., & Wu, J. (2017). Inventory-transportation integrated optimization for maintenance spare parts of high-speed trains. *PloS One, 12*(5), e0176961-e0176961.

Mabini, M.C. (2016). A review of recent literature on repairable-item inventory systems. *Acta Manilana, 64*, 67-74.

Maleyeff, J., Xu. J., Kaffaty, V., and Yang Y. (2022). Optimal service level for repair part inventory considering the cost of holding other parts. Proceedings of the 53rd Annual Conference of the Decision Sciences Institute, 722-734.

Mamer, J.W. & Smith, S.A. (1985). Job completion-based inventory systems: Optimal policies for repair kits and spare machines. *Management Science, 31*(6), 703-718.

Michalak, K. & Knowles, J.D. (2016). Simheuristics for the multiobjective nondeterministic firefighter problem in a time-constrained setting. In Squillero G. & Burelli P. (Eds.),

Applications of Evolutionary Computation: 19th European Conference, Springer International Publishing, 248-265.
Prak, D., Saccani, N., Syntetos, A., Teunter, R., & Visintin, F. (2017). The repair kit problem with positive replenishment lead times and fixed ordering costs. *European Journal of Operational Research, 261*(3), 893-902.
Rabe, M., Deininger, M., & Juan, A.A. (2020). Speeding up computational times in simheuristics combining genetic algorithms with discrete-Event simulation. *Simulation Modelling Practice and Theory, 103*, 102089.
Schuh, P., Schneider, D., Funke, L., & Tracht, K. (2015). Cost-optimal spare parts inventory planning for wind energy systems. *Logistics Research, 8*(1), 1-8.
Seiringer, W., Castaneda, J., Altendorfer, K., Panadero, J., & Juan, A.A. (2022). Applying simheuristics to minimize overall costs of an MRP planned production system. *Algorithms, 15*(2), 40.
Silver, E.A., Pyke, D.A., & Thomas, D.D. (2016). *Inventory and Production Management in Supply Chains*. CRC Press eBooks.
Slay, F.M. &Sherbrooke, C. (1988). The nature of the aircraft component failure process. Technical Report IR701R1, Logistics Management Institute, Washington D.C.
Sun, B., Liu, J., Hao, J., Shen, X., Mao, X., & Song X. (2020). Maintenance decision-making of an urban rail transit system in a regionalized network-wide perspective. *Sustainability (Basel, Switzerland), 12*, 9734.
Teunter, R.H., Babai, M.Z., & Syntetos, A.A. (2010). ABC classification: Service levels and inventory costs. *Production and Operations Management, 19*(3), 343-352.
Welch, P.D. (1983). The statistical analysis of simulation results. In S. S. Lavenberg (Ed.), *The computer performance modeling handbook*. New York: Academic Press.

DECISION SCIENCES INSTITUTE

A Two-Stage Stochastic Programming Model to Plan for the Central Texas Food Bank's Inventory Management after COVID-19

Jakir Hassan
Texas State University - Ingram School of Engineering
San Marcos, TX
Email: oph16@txstate.edu

Clara M. Novoa
Texas State University - Ingram School of Engineering
San Marcos, TX
Email: cn17@txstate.edu

ABSTRACT

The Central Texas Food Bank experienced notable inventory reductions for multiple food categories because of canceled USDA shipments and decreased retail donations after COVID-19. To counteract these supply chain disruptions, this paper presents a two-stage stochastic multi-period inventory model to recommend an optimal mix of food categories and ordering amounts to satisfy uncertain demands at minimum cost. The stochastic model involves a first-stage decision, buying food in each category under uncertain demand, and second-stage decisions or recourse actions to minimize the impact of food scarcity and waste. The two-stage stochastic model's cost-effectiveness is demonstrated through comparison to a deterministic model.

KEYWORDS: Inventory models, Food banks, Supply chain disruptions, Stochastic programming, Uncertain demand

INTRODUCTION

The COVID-19 pandemic impacted the world, and food banks were no exception. Following the COVID-19 pandemic, food banks worldwide faced unprecedented challenges in meeting the increased demand for food assistance. To illustrate the role played by the charitable food sector in addressing the increased levels of food insecurity, Feeding America (2021a) estimated that a minimum of 60 million individuals, constituting one-fifth of the U.S. population, sought aid from food banks, food pantries and similar private food assistance programs during 2020. There was a 50%-55% surge in the total number of individuals seeking support from charitable food assistance compared to 2019 (*Feeding America*, 2021a; Feeding America, 2021b). Similarly, the Feeding Texas network, which comprises 21 food banks representing diverse communities throughout the state of Texas, encountered a surge in demand alongside challenges related to disrupted supply chains and limited availability of volunteers. According to the report by Feeding Texas (2020), one out of every four Texas families experienced food insecurity difficulties during the pandemic. Food insecurity increased from 13% of households in 2019 to 31% in July 2020. Furthermore, Feeding Texas (2020) indicated that from March to August, Texas food banks distributed over 400 million pounds of food to their communities, showing a notable 60% rise compared to the corresponding period in 2019.

The Central Texas Food Bank (CTFB) with headquarters in Austin is the largest hunger relief charity in Central Texas and it has operated or nearly 40 years (Central Texas Food Bank, 2023). In 2020, the CTFB began to experience inventory shortages across its 32 food categories due to COVID-19 supply chain disruptions. Servicing over 300 partner agencies across 21 central Texas counties, these inventory shortages put both CTFB and its constituents in jeopardy. According to CTFB, most partner agencies (81%) expressed that their capacity to assist the community in need would be at risk without the support the CTFB provided. As the pandemic recedes, the CTFB continues facing Inventory Management (IM) issues to ensure efficient and effective food distribution.

IM is a complex problem, particularly for the CTFB, which relies on donations, experiences fluctuating demands, operates under resource constraints such as the perishability of food items and limited storage space, and needs to ensure enough food is on hand to distribute about 64 million pounds of food yearly. Given these challenges, this study aims to apply an operations research model to help the CTFB to address disruptions in supplies and demands post-COVID-19 by accomplishing three objectives: (1) recommending an optimal mix of product categories and quantities to procure over time, (2) achieving effectiveness in procuring the food needed to satisfy the uncertain demands and (3) efficiently minimizing the acquisition and inventory costs.

Figure 1: CTFB supply chain problem researched

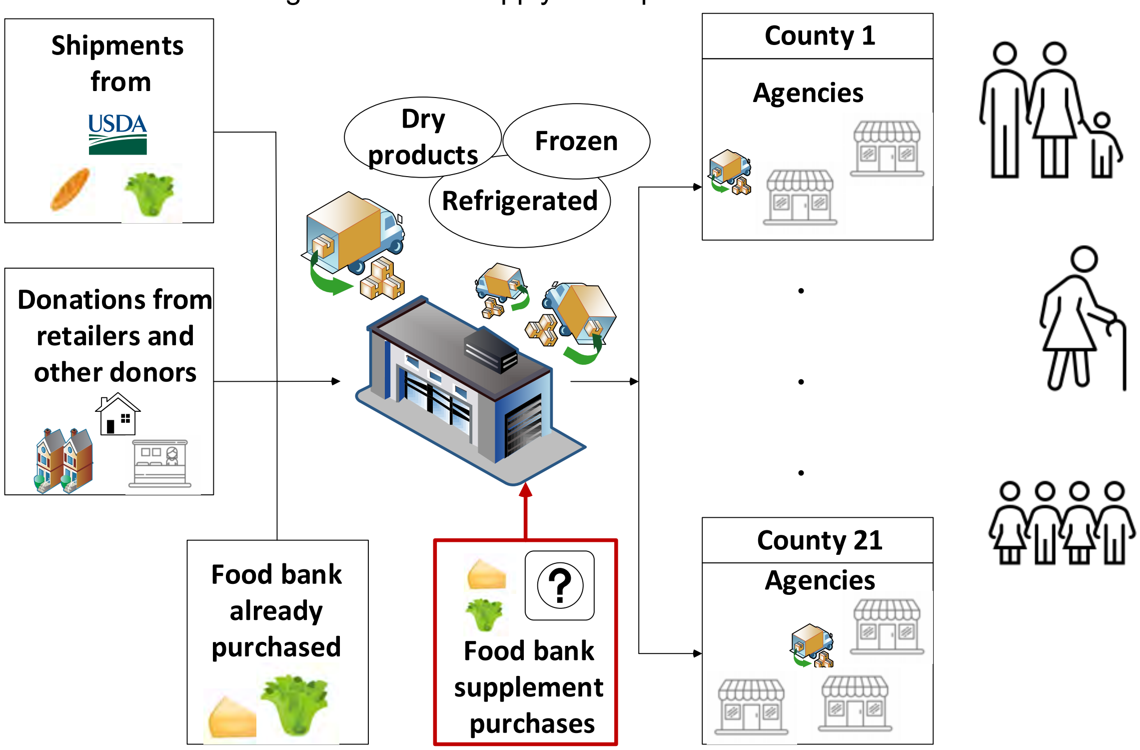


Figure 1 illustrates the CTFB supply chain problem researched in this paper. CTFB gets food from USDA, retailers, and other donors. Besides, CTFB uses some of the donated money to purchase food to cover fluctuations in food donations and satisfy the demands of the community in need. The dry, frozen, and cooler or refrigerated food received is stored in the CTFB warehouse and distributed to 300 partner agencies, such as food pantries and shelters in 21 counties, that supply the food to their final customers. However, because of uncertainty in demands and reductions in donations CTFB ends up needing to make supplemental purchases to satisfy the demands. The research question investigated in this paper is if it is feasible to find an optimal monthly procurement plan for CTFB supplemental purchases in a post-COVID-19 environment given CTFB's financial and resource limitations.

This paper proposes a two-stage stochastic programming (TSSP) model to address the research question above. By incorporating uncertainty in food demand and CTFB supply chain constraints, such as limited storage capacity, waste limits of perishable items, and budget available, the proposed model aims to minimize costs associated with sourcing food, storing inventory, and disposing of perished food, and enhance the CTFB's ability to provide essential food assistance. A TSSP model is a robust mathematical framework because it simultaneously plans for the initial decision-making stage and the subsequent operational stage accounting for uncertainties represented in multiple scenarios and allowing for adaptability and flexibility in decision-making.

This study draws upon the existing research on stochastic programming models for IM and supply chain optimization. Several notable contributions have explored the application of stochastic programming in various domains, such as agriculture, manufacturing, and logistics. For instance, Nguyen & Chen (2019) introduced a multi-period inventory model for a perishable product with stochastic supply and demand in a rolling horizon framework. The main objective was to minimize the expected total cost of ordering, purchasing, holding, shortage, and waste. Additionally, Gupta & Maranas (2000) proposed a TSSP approach for incorporating demand uncertainty in multisite midterm supply-chain planning problems. In this paper, we build upon these foundations and adopt the TSSP approach to address the unique challenges the CTFB faces. By applying a TSSP approach, this study contributes to the scarce body of knowledge in IM for food banks using TSSP and provides actionable insights for food bank operations. Furthermore, by optimizing the inventory system and enhancing operational efficiency, the proposed model assists the CTFB in its mission to alleviate hunger and nourish the community in Central Texas.

The remainder of this paper is organized as follows. Section 2 reviews relevant literature on food banks' IM and on stochastic programming models proposed to model uncertain demand and supply for food banks or other uncertainties in similar contexts. Section 3 presents the TSSP model proposed and a deterministic model developed for comparison purposes. Section 4 describes the case study. Section 5 provides the numerical results and discussion. Section 6 presents the sensitivity analysis. Section 7 is the conclusions section, which summarizes the findings, highlights the practical insights, and suggests future research directions.

LITERATURE REVIEW

Literature Review on the Effect of COVID-19 on Food Bank Operations

In 2023, a report from the Food Security Information Network and Global Network Against Food Crises (FSIN and Global Network Against Food Crises, 2023), reported that 258 million people in 58 countries were affected by food insecurity in 2022. The COVID-19 effects were one of the contributors to this issue in various regions worldwide. Esmaeilidouki et al. (2021) mentioned that food insecurity is on the rise worldwide, affecting over one billion people who lack access to adequate nutrition. As non-profit organizations, food banks have prioritized helping food-insecure individuals by delivering food donations. The significance of these humanitarian organizations is demonstrated by the increasing number of food-insecure households and individuals served by food banks in affluent nations. Each year, a growing number of people are served by food banks, which play a vital role in eliminating food shortages. Besides, during the COVID-19 pandemic, governments implemented new policies and funding programs to aid food banks in meeting the food security needs of the community. For example, the United States Department of Agriculture (USDA) initiated the "Farmers to Families Food Box Program," in which food banks such as Los Angeles Regional Food Bank (LRFB) served as intermediaries between suppliers and agencies to coordinate supply and demand (Blackmon et al., 2021). The LRFB, Salesforce, and the University of California Los Angeles developed a decision support system in 45 days to help the LRFB to function as an efficient intermediary.

Food safety, user perception, food insecurity, and food bank operations are the four primary subject areas highlighted by current research on food bank operations and food insecurity. For example, a detailed literature review on food banks and their operational challenges under COVID-19 disruption was conducted by Esmaeilidouki et al. (2021) based on the literature published between 2000 to 2021. Esmaeilidouki et al. (2021) used an effective three-stage process in which 48 relevant papers were selected and analyzed after initially searching 200 papers.

Literature Review on Food Bank Operations

Another noteworthy literature review on food banks is McIntyre et al. (2016), who evaluated 33 articles on food bank operations from 1998 to 2014 to identify if the operational challenges faced by food banks suggest that these organizations may be perpetuating inequity. On the other hand, Ataseven et al. (2018) used an online survey methodology targeting 202 food banks to test that organizational and human assets drive supply chain integration (i.e., supply integration, internal integration, and demand integration) and are critical for sustained operation of food banks. The analysis of the responses of 100 participating food banks demonstrated the validity of the hypothesis.

Overall, the literature regarding food bank operations is extensive. Therefore, Esmaeilidouki et al. (2021) have classified it into four areas: 1) Distribution Management, 2) Facility Planning, 3) Volunteers Scheduling and 4) Inventory Management (IM). This paper concentrates on IM of the food bank supply chain. However, literature on food bank IM is limited and IM problems become significantly more challenging when faced with uncertain demand. Publications in Kim et al. (2015) and Solyalı et al. (2015) discuss the difficulty of addressing IM under uncertain demand.

Some relevant aspects of food bank IM problems are variations in donations, changes in demands, agencies' uncertain supply (Orgut et al., 2017; Orgut et al., 2018), optimal ways to allocate extra receiving capacity to counties (Orgut et al., 2016), and management of perishable food (Giuseppe et al., 2014; Mandal et al., 2021). Orgut et al. (2016) provided a deterministic network-flow model to minimize the amount of undistributed food and keep equitable distribution to each county. An extension of this model was presented to determine optimal ways of allocating new extra receiving capacity to bottleneck counties. Furthermore, Orgut et al. (2017) considered stochastic receiving capacities at the agencies, affected by budget and workforce availability, to ensure equitable distribution of food donations in a single-period two-stage stochastic model. Then, Orgut et al. (2018) extended the work of Orgut et al. (2016) by implementing two robust optimization models to achieve adequate distribution of food donations considering uncertainty in counties and agencies' receiving capacities. In their study, food demands remain as a deterministic parameter.

Regarding the integration of distribution management, facility planning, and IM for food bank operations, Martins et al. (2019) presented a mixed integer programming model to redesign a food bank supply chain network. Their work deals with accounting for three objectives (economic, environmental, and social aspects of sustainability) and their associated tradeoffs. The numerical study uses problem instances from the Portuguese Federation of Food Banks (FPBA) network.

Literature Review on Modeling Supply and Demand Uncertainties in Food Bank Operations with Two-Stage Stochastic Programs

There are a few contributions in the food bank literature where supply and demand uncertainties are modeled in a two-stage stochastic model. For instance, Kaviyani-Charati et al. (2022) developed a comprehensive two-stage stochastic model for a non-profit food bank supply chain that incorporates multiple objectives, stages, products, and periods. The model considered the three pillars of sustainability (economic, social, and environmental), as well as time constraints and a diverse transportation fleet while aiming at multiple objectives such as reducing food waste, facilitating the collection and reusing of surplus food, minimizing environmental and economic impacts, enhancing food security, and fulfilling demand. In addition, this contribution modeled uncertainty on the amount of monetary donation, unmet demand, and extra warehouse and transportation capacity.

Tofighi et al. (2016) proposed a two-stage scenario-based possibilistic-stochastic programming (SBPSP) approach to design a logistic network for disaster relief operations under mixed uncertainty. In this case, both the supply and demand were considered uncertain. In a different context, Dillon et al. (2017), proposed a two-stage stochastic programming model to manage the inventory of red blood cells and minimize operational expenses, blood shortage and wastage caused by expiration while considering perishability and demand uncertainty.

Literature Review on Inventory Management (IM) Considering Perishability of Products and Food Waste

Another significant issue related to IM is the perishability of products or food wastage, and the food bank's inventory is no exception in this case. According to Mallidis et al., (2022), food security and availability are two ethical and moral concerns related to food waste's social consequences. Mourad (2016) and Papargyropoulou et al. (2014), state that prevention efforts (e.g., reducing surplus at the source, optimizing processes) have priority over re-use for human consumption through redistribution networks and welfare organizations (including food banks and charities). Nguyen & Chen (2019) optimized a perishable inventory system with the main objective of minimizing the expected total cost of ordering, purchasing, inventory holding, shortage, and waste.

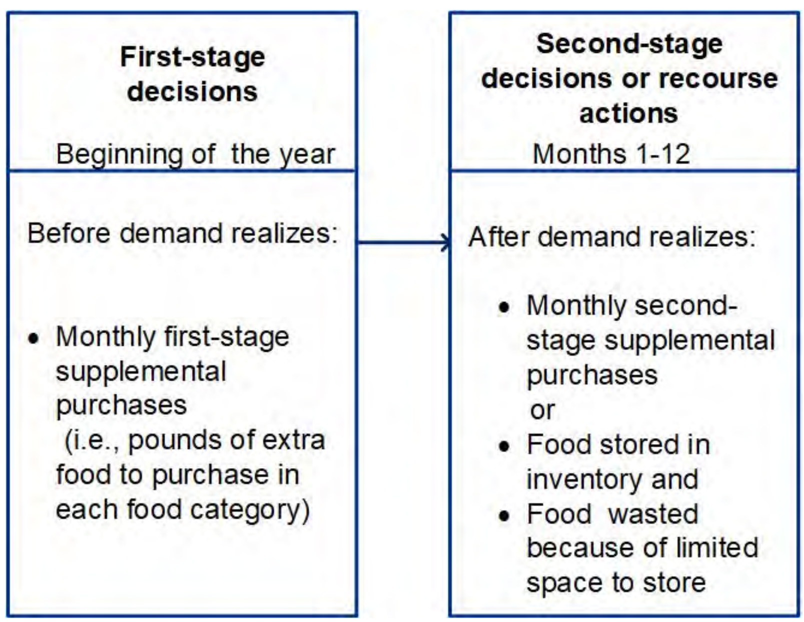
Besides, a food bank-related study was conducted by Mandal et al. (2021), where a decision support system was proposed for an aggregator that connects retailers to food banks, and it may reduce food wastage while maximizing profit and minimizing the environmental impact. Although our research work disseminated in this paper has not considered the perishability for some of the food categories, we have included the cost of food wastage in the objective function that minimizes the overall operational cost.

In summary, this paper includes several main contributions. Firstly, we formulate a two-stage, multi-period, multi-product stochastic programming IM considering stochastic demands from the agencies to minimize operational costs, including acquisition, inventory and wasted food costs. Secondly, we recommend an optimal mix of food categories to satisfy demand and mitigate disruptions in supply. Thirdly, we help decision-makers by demonstrating the cost-effectiveness of using our stochastic model instead of a deterministic model.

TSSP MODEL FORMULATION

The CTFB IM problem is formulated as a multi-period, multi-product TSSP. As mentioned in the introduction and presented in Figure 1, CTFB used some of the donated money to purchase food to cover fluctuations in food donations and be prepared to satisfy the increasing demands of the community in need. Thus, the TSSP considers the amounts of food already purchased and the uncertainty of final demand for each food category to determine an optimal strategy to perform supplemental purchases or other needed recourse actions that satisfy the demand and other realistic constraints. Then, the TSSP model (TSSPM) determines the additional amounts of food to purchase before realizing the demand uncertainty (first-stage decision), and after realizing it (second-stage decision) and additional recourse actions to deal with the uncertainty in demand, such as storing in inventory or removing food that cannot be stored in the warehouse due to space limitations. Figure 2 illustrates the decisions of the model in each of the two stages.

Figure 2: Decision variables in the CTFB TSSP model



Tables 1-3 present the notation for indexes, sets, parameters, and decision variables in the TSSP model, and the TSSP formulation is immediately below the tables. In the TSSP model, equation (1) is the objective function that includes the costs of food received from USDA, original CTFB purchases, first-stage and second-stage supplemental purchases, food wasted (i.e., food that has to be removed from the warehouse due to space limitations), and carrying inventory. Constraints (2) and (3) are to conserve the product flow and ensure that the total amount of available food equals the sum of the food demanded, sent to inventory, and wasted. Constraint (4) assures that the amount of wasted food is less than the first-stage supplemental purchases plus the original purchases and all donations received. Constraint (5) defines that the amount of food the inventory carries cannot exceed storage capacity. Constraint (6) assures the overall monthly cost of procuring the food does not exceed the monthly budget. Then, the sign constraints (7) and (8) assure that first and second-stage supplemental purchases, amount of food wasted, and inventory carried are non-negative.

The TSSP model proposed assumes: (1) deterministic supply (i.e., USDA shipments, donations, and already completed purchases by food bank) (2) no perishability of food categories (3) no exchange of food between food banks (4) stochastic demand scenarios coming from historical data collected (2020 and 2021).

Table 1: TSSP Model Sets and Indexes

Notation	Description
F	Set of food categories included in the model, $F = \{4,6,7,10,14,15,16,23,27,28\}$ and index $f \in F$
T	Set of time periods $T=\{1,2,\ldots,12\}$ and index $t \in T$
S	Set of scenarios $S = \{2020, 2021\}$. The scenarios contain historical demand data for the years 2020 and 2021 and index $s \in S$
C	Set of types of storage types $C = \{1,2,3\}$ in the CTFB warehouse; 1 is for dry, 2 is for cooler or refrigerated, and 3 is for frozen, and index $j \in C$
D	Set of dry food categories included in the model, $D = \{4,6,10,14,16,23,27\}$ and index $d \in D$
A	Set of cooler or refrigerated food categories included in the model, $A = \{7,28\}$ and index $a \in A$
R	Set of frozen food categories included in the model, $R = \{15\}$ and index $r \in R$

Table 2: TSSP Model Parameters

Notation	Description
r_{ft}	Cost per pound of handling and distributing food category f from USDA at period t.
e_{ft}	Pounds of food category f handled and distributed by USDA at period t.
n_t	Total cost of original purchases for food at period t.
o_{ft}	Pounds originally purchased for food category f at period t.
q_{ft}	Cost per pound of first-stage supplemental purchases for food category f at period t.
u_{ft}	Cost per pound of second-stage supplemental purchases for food category f at period t.
h_d	Inventory carrying cost per pound for dry food d
h_a	Inventory carrying cost per pound for cooler food a
h_r	Inventory carrying cost per pound for frozen food r
l_{ft}	Pounds of food category f donated by other donors and retailers at period t.
p_s	Probability of occurrence for scenario s ($\sum_{s \in S} p_s = 1$)
d_{fts}	Pounds demanded of food category f at period t under scenario s.
g_j	Pounds of storage capacity type j.
b_t	Monthly budget of CTFB at period t.

Table 3: TSSP Model Decision Variables	
Notation	Description
x_{ft}	Pounds of first-stage supplemental purchase for each food category f at period t.
y_{fts}	Pounds of second-stage extra purchase for each food category f at period t under scenario s.
m_{fts}	Pounds of food categories f wasted at period t under scenario s
$z_{f(t-1)}$	Pounds of inventory carried for food category f in period $t-1$.
z_{fts}	Pounds of inventory carried for food category f at period t under scenario s.

Model 1. TSSP Model:

$$Min\ z_{TSSPM} = \sum_{t\in T} n_t + \sum_{f\in F}\sum_{t\in T}\left(r_{ft}e_{ft} + q_{ft}x_{ft}\right) + \sum_{s\in S} p_s \left[\sum_{f\in f}\sum_{t\in T}\left(u_{ft}y_{fts} + q_{ft}m_{fts}\right)\right]$$
$$+ \sum_{s\in S} p_s \left[\sum_{t\in T}\sum_{d\in D} h_d z_{dts} + \sum_{t\in T}\sum_{a\in A} h_a z_{ats} + \sum_{t\in T}\sum_{r\in R} h_r z_{rts}\right] \quad (1)$$

s.t.

$$z_{f(t-1)} + e_{ft} + o_{ft} + l_{ft} + x_{ft} + y_{fts} = d_{fts} + z_{fts} + m_{fts}\,, \qquad \forall f \in F, t = 1, \forall s \in S \quad (2)$$

$$z_{f(t-1)s} + e_{ft} + o_{ft} + l_{ft} + x_{ft} + y_{fts} = d_{fts} + z_{fts} + m_{fts}\,, \qquad \forall f \in F, \forall t \in T\backslash\{1\}, \forall s \in S \quad (3)$$

$$m_{fts} \le e_{ft} + o_{ft} + l_{ft} + x_{ft}\,, \qquad \forall f \in F, \forall t \in T, \forall s \in S \quad (4)$$

$$\sum_{f\in c_j} z_{fts} \le g_j\,, \qquad \forall j \in c, \forall t \in T, \forall s \in S \quad (5)$$

$$\sum_{f\in F}\left(r_{ft}e_{ft} + q_{ft}x_{ft} + u_{ft}y_{fts}\right) + n_t \le b_t\,, \qquad \forall t \in T, \forall s \in S \quad (6)$$

$$x_{ft}, y_{fts}, m_{fts}, z_{fts} \ \ge 0\,, \qquad \forall f \in F, \forall t \in T, \forall s \in S \quad (7)$$

$$z_{f(t-1)} \ \ge 0\,, \qquad \forall f \in F, t = 1 \quad (8)$$

To assess the performance of the proposed TSSP model (Model 1) for solving the CTFB supply chain problem, the following deterministic model (Model 2) is formulated. Some notations in Tables 1-3 and the notation in Table 4 are used in the deterministic model.

Table 4: Additional Decision Variables (D.V.) and Parameters (Par.) for the Deterministic Model		
Notation	Category	Description
$\widehat{d_{ft}}$	Par.	Estimated mean pounds demanded for food category f at period t
m_{ft}	D.V.	Pounds of food categories f wasted at period t
$z_{f(t-1)}$	D.V.	Pounds of inventory carried for food category f in period $t-1$.
z_{ft}	D.V.	Pounds of inventory carried for food category f at period t.

Equations 1-2 in the deterministic model have the same purpose as the ones in the TSSP model. Similarly, Equations 3-5 in the deterministic model have a similar meaning to Equations 4-6 in the stochastic model. However, in the deterministic model, the scenarios are eliminated because only the mean demand for each food category is included in the model.

Model 2. Deterministic Model:

$$Min\ z_{Det} = \sum_{t\in T} n_t + \sum_{f\in F}\sum_{t\in T}\left(r_{ft}e_{ft} + q_{ft}x_{ft} + q_{ft}m_{ft}\right) + \sum_{t\in T}\sum_{d\in D} h_d z_{dt} + \sum_{t\in T}\sum_{a\in A} h_a z_{at} + \sum_{t\in T}\sum_{r\in R} h_r z_{rt} \tag{1}$$

s.t.

$$z_{f(t-1)} + e_{ft} + o_{ft} + l_{ft} + x_{ft} = \widehat{d_{ft}} + z_{ft} + m_{ft}\,, \qquad \forall f \in F, t \in T \tag{2}$$

$$m_{ft} \leq e_{ft} + o_{ft} + l_{ft} + x_{ft}\ , \qquad \forall f \in F, \forall t \in T \tag{3}$$

$$\sum_{f\in c_j} z_{ft} \leq g_j\ , \ \ \forall j \in c, \forall t \in T \tag{4}$$

$$\sum_{f\in F}\left(r_{ft}e_{ft} + q_{ft}x_{ft}\right) + n_t \leq b_t\,, \qquad \forall t \in T \tag{5}$$

$$x_{ft}, m_{ft}, z_{ft} \ \geq 0\,, \qquad \forall f \in F, \forall t \in T \tag{6}$$

CASE STUDY

The case study is based on data provided by the CTFB for the years 2019, 2020, and 2021. After analyzing Pareto Charts for the pounds of food distributed for each of the 32 food categories stored in the CTFB each month during the years 2019-2021, 10 food categories were consistently selected as the most demanded ones. Therefore, only these 10 categories were included in the TSSP model. Figure 3 presents a consolidated Pareto Chart for the year 2020 showing the most relevant food categories. Table 5 lists the names of the ten food categories and the percentage that each one contributed to the total demand in 2020. Note that the percentages of contribution for these 10 categories account for about 90% of the pounds of food distributed by the CTFB and that the percentages of contribution per category are relatively similar for the years 2019-2021.

Table 6 provides the storage capacities for each of the types of storage (dry, cooler or refrigerated and frozen) used in the CTFB warehouse. As mentioned in Table 1, the model considers only scenarios corresponding to the monthly demands for the years 2020 and 2021. The demands for the year 2019 was excluded because these demands were significantly different from the post-COVID years 2020 and 2021. CTFB provided the initial inventory and the purchase cost and extra purchase cost per pound in each food category. Besides monthly data for demands, CTFB also provided pounds of food donated by USDA, pounds of food already purchased by CTFB, and pounds of food donated by other donors or retailers in each month and food category. The monthly budget is about 7.2 million.

Figure 3: Pounds distributed by food category in 2020

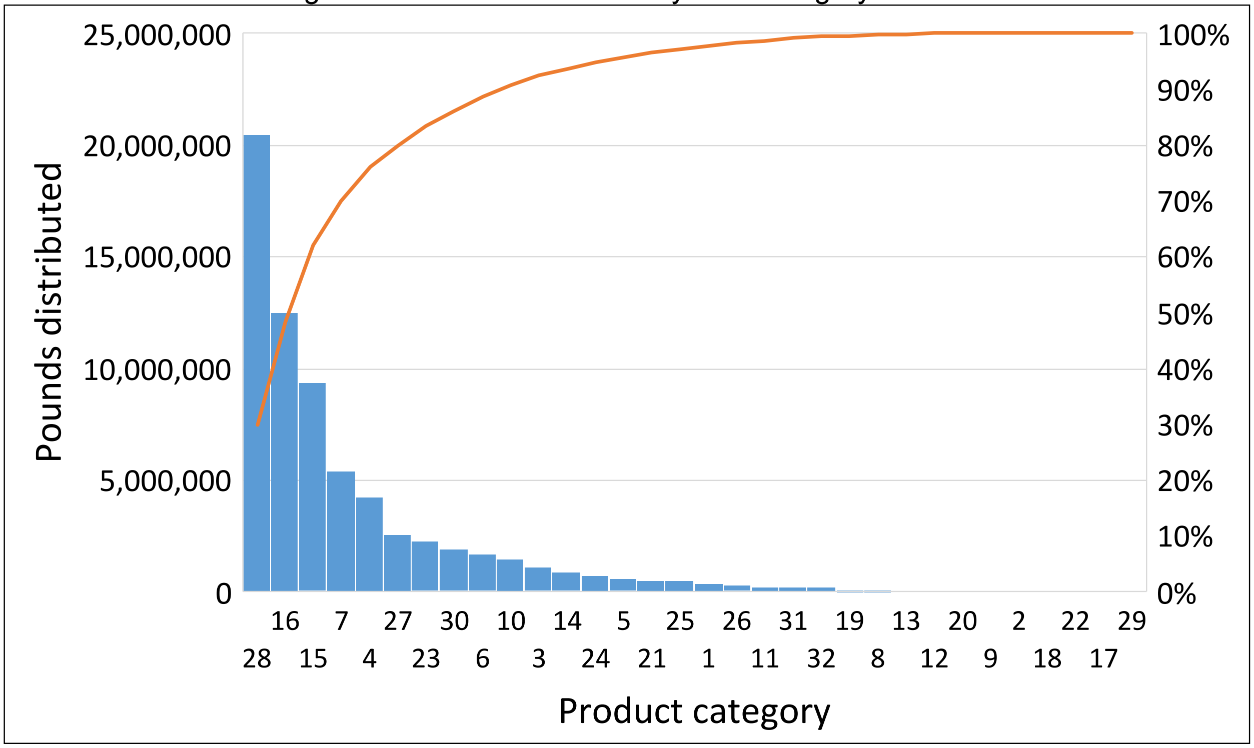


Table 5: The 10 Food Categories Included in the TSSP Model					
Category No.	Description	Storage Type	2019	2020	2021
28	Fresh fruit & vegetables	Cooler	41%	30%	26%
16	Mixed & assorted	Dry	7%	18%	24%
15	Meat, fish, poultry	Frozen	10%	14%	10%
7	Dairy	Cooler	5%	8%	7%
4	Bread and bakery	Dry	8%	6%	7%
27	Vegetables canned	Dry	5%	4%	4%
23	Protein non-meat	Dry	4%	3%	3%
6	Complete meal entrée soup	Dry	2%	2%	2%
10	Fruit canned	Dry	3%	2%	2%
14	Juice	Dry	2%	1%	3%
Total			82%	88%	85%

Table 6: Storage Capacity for Each Storage Type (pounds)	
Storage Type	Capacity
Dry	5,224,800
Cooler or refrigerated	2,226,000
Frozen	2,730,000

NUMERICAL RESULTS AND DISCUSSION

The TSSP model was coded in AMPL using AMPL IDE and solved with the CPLEX solver. All experiments were conducted on a PC with a 1.30 GHz 12 Gen Intel® Core™ i5-1235U processor, 12 GB RAM, and running Windows 11 (64-bit).

The customary three-step process to compare deterministic vs. two-stage stochastic model solutions was adopted. First, the deterministic model and the TSSP models were solved to determine their optimal solutions, z_{TSSPM} and z_{Det} Then, the supplemental purchases,x_{ft}, prescribed by the deterministic model, were input as parameters into the TSSP model and the stochastic model was solved to find only the values of the second-stage variables and a new objective function cost notated as $z_{Det\_into\_TSSPM}$. The cost saving (i.e., a cost advantage) from using the TSSP model is assessed as $z_{Det\_into\_TSSPM} - z_{TSSPM}$. The assessed cost saving was $10,060,900 ($40,535,000 - $30,474,100). It is an annual cost reduction of about 25%.

Figures 4 and 5 compare the first-stage and second-stage supplemental purchases per food category found by the TSSP model and the model that plugged the deterministic solution into the stochastic model, respectively. Figure 4 presents the first-stage supplemental purchases (blue bars) and the second-stage extra purchases found by the TSSP model for the demand scenario year 2020 (orange bars) and for the demand scenario year 2021(grey bars). On the other hand, Figure 5 presents the deterministic solution for first-stage supplemental purchases (blue bars) and the second-stage extra purchases needed under demand scenario year 2020 (orange bars) and demand scenario year 2021 (grey bars). A significant difference between the two graphs is that the TSSP model emphasizes second-stage purchases on several food categories, while the deterministic solution ensures that larger first-stage purchases occur for most of the food categories, but plugging these purchases into the stochastic model ultimately triggers wastage of food incurring in extra cost.

Both models decide not to buy bread and bakery at any stage and it is because donations of bread and bakery exceed the demands. Dairy products are consistently purchased by both models, mostly in the first stage because even if it is a product category with high demand and high variability (i.e., high standard deviation for the demand) it has relatively low price. Juice is consistently purchased in the second-stage only because of its low demand variability and its relatively low price. The models differ on the way they decide to purchase for the mixed & assorted category which is one of the categories with the highest variability in monthly demands. Because the deterministic model observes only the mean demands, it purchases mostly in the first-stage. On the other hand, the stochastic model postpones the supplemental purchases of this highly demanded food category to the second stage. The models consistently postpone the supplemental purchases of fresh fruit & vegetables to the second stage. This is a consistent decision explained by the high demands, high demand variability and space constraints for cooler or refrigerated food.

Figure 4. Thousands of food purchased by the TSSP model

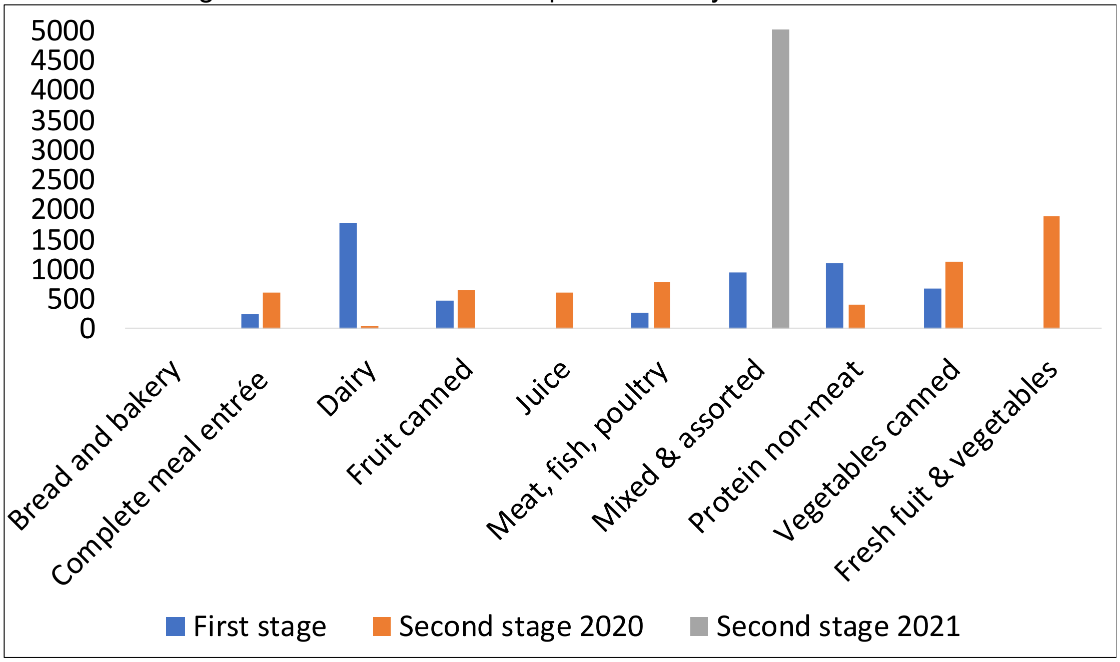
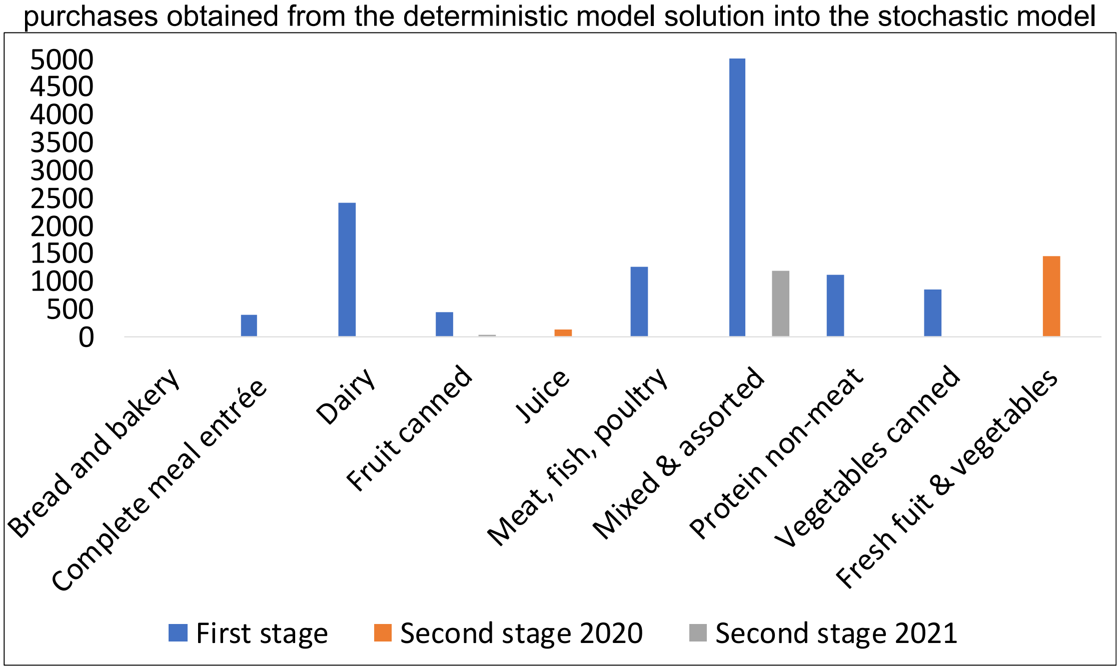


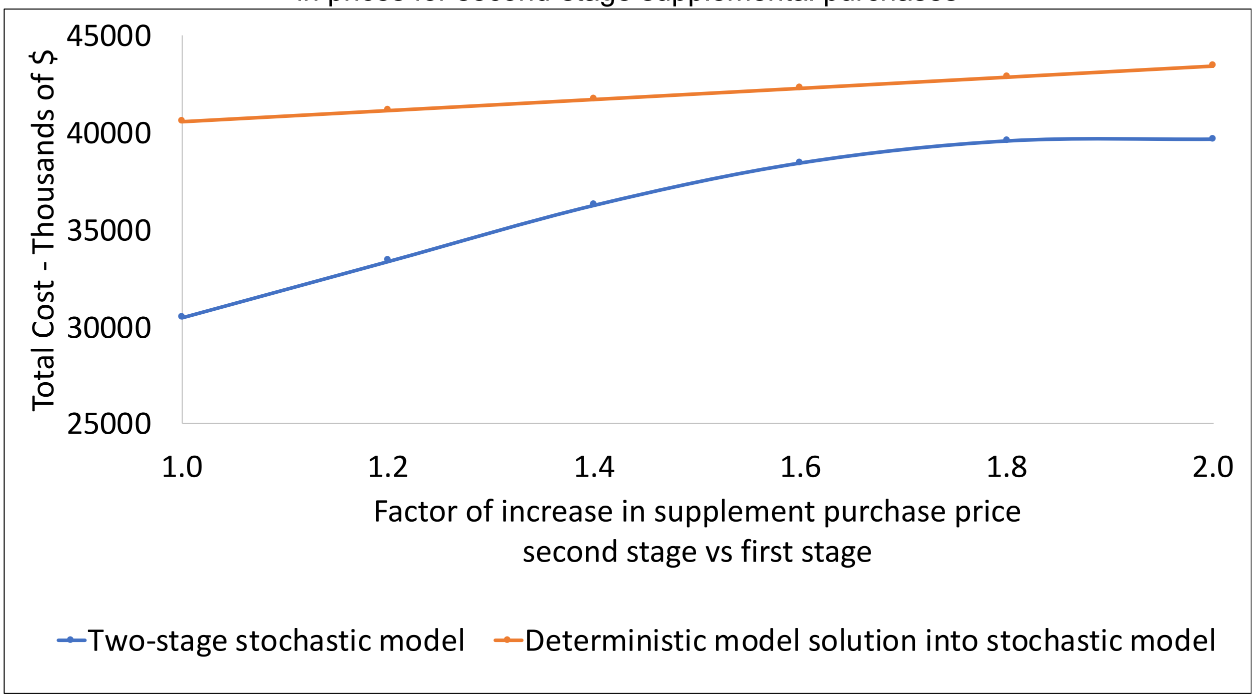
Figure 5. Thousands of food purchased by the model that plugs supplemental purchases obtained from the deterministic model solution into the stochastic model



SENSITIVITY ANALYSIS

A sensitivity analysis was performed with the models to explore the effect of changes in the supplemental purchase costs on the models' total cost. For the base case presented in the previous section, the first-stage supplemental purchase cost, q_{ft}, is considered equal to the second-stage supplemental purchase cost, u_{ft} (i.e., $\frac{u_{ft}}{q_{ft}} = 1$). In the sensitivity analysis, the second-stage supplemental purchase cost is increased vs. the first-stage supplemental purchase cost by 20%, 40%, 60%, 80% and 100% (i.e., u_{ft} is 1.2, 1.4, 1.6, 1.8, and 2.0 times larger than q_{ft}). Figure 5 illustrates that the cost advantage between the TSSP model and the model that plugs the deterministic solution in the stochastic model decreases. Thus, the highest total cost difference is observed when $q_{ft} = u_{ft}$. This result agrees with the author's expectations. An excessive increase in the second-stage extra purchase cost affects the cost advantage of postponing the purchase of food categories with high demand variability offered by the TSSP model.

Figure 6: Sensitivity analysis for total cost vs. increases in prices for second-stage supplemental purchases



CONCLUSION

This paper addressed a critical inventory management issue, procurement plans for the Central Texas Food Bank in the aftermath of the COVID-19 pandemic. This pandemic has caused supply chain disruptions such as increased demand, reduced donations, and cancelation of shipments for CTFB, which ultimately resulted in inventory shortages in some food categories. This paper concentrated on supply chain disruption challenges, specifically on the rises in demand for various food categories considering them as stochastic parameters. The research study set clear objectives such as recommending an optimal mix of food categories based on scenarios created from the historical demand data, identifying effective acquisition of food amounts to satisfy the demand of a community in need, and minimizing the costs related to procurement, inventory, and food wasted. The findings demonstrate that the inherent uncertainties and risks associated with food banks can be effectively captured and managed through two-stage stochastic programming.

Considering the uncertainty in demand, the two-stage stochastic programming model provides decision-makers at the Central Texas Food Bank with more robust inventory planning strategies. The paper explored the benefits of a stochastic approach compared to a deterministic model through sensitivity analysis. The primary focus was determining the appropriate quantities of food to procure for each category before the actual demand uncertainty was realized (i.e., the first-stage decisions). Subsequently, second-stage decisions or recourse actions were found to supplement the initial purchases. The two-stage model enables decision-makers to make proactive decisions in the first stage based on available information and then revise those decisions in the second stage as additional information becomes available. This two-stage framework allows for greater responsiveness and agility in inventory planning, further enhancing cost efficiency. A systematic comparative approach was used to assess the cost difference between the stochastic and deterministic models presented.

The results analysis reveals a clear cost advantage in utilizing the two-stage stochastic programming model for inventory planning. The model is an effective tool for solving the procurement issues CTFB has experienced after COVID-19. Based on the results of this study, some future research aspects were also found which can be included in the model, such as variability in donations, increasing the number of scenarios by sampling from forecasted demands, and incorporating optimization of the food distribution to agencies.

ACKNOLEDGEMENTS

The authors are grateful to the Central Texas Food Bank (CTFB) in Austin, TX. CTFB provided the data and valuable input for this project. The cooperation of the all the staff involved with managing and improving the CTFB was fundamental for the completion of this research.

REFERENCES

Ataseven, C., Nair, A., & Ferguson, M. (2018). An examination of the relationship between intellectual capital and supply chain integration in humanitarian aid organizations: A survey-based investigation of food banks. *Decision Sciences*, 49, 827-862. https://doi.org/10.1111/deci.12300

Blackmon, L., Chan, R., Carbral, O., Chintapally, G., Dhara, S., Felix, P., Jagdish, A., Konakalla,S., Labana, J., McIlvain, J., Stone, J., Tang, C. S., Torres, J., & Wu, W. (2021). Rapid development of a decision support system to alleviate food insecurity at the Los Angeles regional food bank amid the COVID-19 pandemic. *Production and Operations Management*, 30(10), 3391-3407.

Central Texas Food Bank. (2023). About us. Retrieved from https://www.centraltexasfoodbank.org/ , January 2023.

Dillon, M., Oliveira, F., & Abbasi, B. (2017). A two-stage stochastic programming model for inventory management in the blood supply chain. *International Journal of Production Economics*, 187, 27-41. https://doi.org/10.1016/j.ijpe.2017.02.006

Esmaeilidouki, A., Rambe, M., Ardestani-Jaafari, A., Li, E., & Marcolin, B. (2021). Food bank operations: Review of OR methods and challenges during COVID-19. Retrieved from https://doi.org/10.13140/RG.2.2.17608.90889 , May 15, 2023.

Feeding America. (2021a). Feeding America estimates at least 60 million people turned to charitable food assistance in 2020. Retrieved from https://www.feedingamerica.org/about-us/press-room/60-million-turned-to-charitable-food-2020#:~:text=Feeding%20America%20announced%20today%20new,by%20the%20COVID%2D19%20pandemic. , May 15, 2023.

Feeding America. (2021b). Doing more to end hunger - 2021 Annual report. Retrieved from https://www.feedingamerica.org/sites/default/files/2022-01/FA_2021AnnReport_FULL_012522.pdf , May 15, 2023.

Feeding Texas. (2020). COVID-19 impact - Texas food banking + food insecurity. Retrieved from https://www.feedingtexas.org/wp-content/uploads/2020/11/FT_COVIDReport.pdf , May 15, 2023.

FSIN and Global Network Against Food Crises. (2023). Global report on food crises 2023 - Joint analysis for better decisions. Rome. Retrieved from https://www.fsinplatform.org/sites/default/files/resources/files/GRFC2023-hi-res.pdf , May 31, 2023.

Giuseppe, A., Mario, E., & Cinzia, M. (2014). Economic benefits from food recovery at the retail stage: An application to Italian food chains. *Waste Management*, 34(7), 1306-1316. https://doi.org/10.1016/j.wasman.2014.02.018

Gupta, A., & Maranas, C. D. (2000). A two-stage modeling and solution framework for multisite midterm planning under demand uncertainty. *Industrial & Engineering Chemistry Research*, 39(10), 3799-3813. https://doi.org/10.1021/ie9909284

Kaviyani-Charati, M., Ameli, M., Heidarzadeh Souraki, F., & Jabbarzadeh, A. (2022). Sustainable network design for a non-profit food bank supply chain with a heterogeneous fleet under uncertainty. *Computers & Industrial Engineering*, 171, 108442. https://doi.org/10.1016/j.cie.2022.108442

Kim, G., Wu, K., & Huang, E. (2015). Optimal inventory control in a multi-period newsvendor problem with non-stationary demand. *Advanced Engineering Informatics*, 29(1), 139-145. https://doi.org/10.1016/j.aei.2014.12.002

Mallidis, I., Sariannidis, N., Vlachos, D., Yakavenka, V., Aifadopoulou, G., & Zopounidis, K. (2022). Optimal inventory control policies for avoiding food waste. *Operational Research An International Journal* , 22(1), 685-701. https://doi.org/10.1007/s12351-020-00554-w

Mandal, J., Mitra, R., Gupta, V. K., Subramanian, N., Kayikci, Y., & Tiwari, M. K. (2021). Optimal allocation of near-expiry food in a retailer-foodbank supply network with economic and environmental considerations: An aggregator's perspective. *Journal of Cleaner Production*, 318, 128481. https://doi.org/10.1016/j.jclepro.2021.128481

Martins, C. L., Melo, M. T., & Pato, M. V. (2019). Redesigning a food bank supply chain network in a triple bottom line context. *International Journal of Production Economics*, 214, 234-247.

https://doi.org/10.1016/j.ijpe.2018.11.011

McIntyre, L., Tougas, D., Rondeau, K., & Mah, C. L. (2016). "In"-sights about food banks from acritical interpretive synthesis of the academic literature. *Agriculture and Human Values*, 33(4), 843-859.

Mourad, M. (2016). Recycling, recovering and preventing "food waste": competing solutions for food systems sustainability in the United States and France. *Journal of Cleaner Production*, 126, 461-477. https://doi.org/10.1016/j.jclepro.2016.03.084

Nguyen, & Chen. (2019). Optimization of a perishable inventory system with both stochastic demand and supply: comparison of two scenario approaches. *Croatian Operational Research Review*, 10(1), 175-185. https://doi.org/10.17535/crorr.2019.0016

Orgut, I. S., Ivy, J., Uzsoy, R., & Wilson, J. R. (2016). Modeling for the equitable and effective distribution of donated food under capacity constraints. *IIE Transactions*, 48(3), 252-266. https://doi.org/10.1080/0740817X.2015.1063792

Orgut, I.S., Ivy, J., & Uzsoy, R. (2017). Modeling for the equitable and effective distribution of food donations under stochastic receiving capacities. *IISE Transactions*, 49(6), 567-578. https://doi.org/10.1080/24725854.2017.1300358

Orgut, I.S., Ivy, J. S., Uzsoy, R., & Hale, C. (2018). Robust optimization approaches for the equitable and effective distribution of donated food. *European Journal of Operational Research*, 269(2), 516-531. https://doi.org/https://doi.org/10.1016/j.ejor.2018.02.017

Papargyropoulou, E., Lozano, R., K. Steinberger, J., Wright, N., & Ujang, Z. b. (2014). The food waste hierarchy as a framework for the management of food surplus and food waste. *Journal of Cleaner Production*, 76, 106-115. https://doi.org/10.1016/j.jclepro.2014.04.020

Solyalı, O., Cordeau, J.-F., & Laporte, G. (2015). The Impact of Modeling on Robust Inventory Management Under Demand Uncertainty. *Management Science*, 62(4), 1188-1201. https://doi.org/10.1287/mnsc.2015.2183

Tofighi, S., Torabi, S. A., & Mansouri, S. A. (2016). Humanitarian logistics network design under mixed uncertainty. *European Journal of Operational Research*, 250(1), 239-250. https://doi.org/10.1016/j.ejor.2015.08.059

DECISION SCIENCES INSTITUTE

Additive Manufacturing as a Catalyst for Supply Chain Resilience: An Empirical Analysis

Abdullah Yildizbasi
Worcester Polytechnic Institute
Email: ayildizbasi@wpi.edu

Cihat Ozturk
Ankara Yildirim Beyazit University
Email: cozturk@aybu.edu.tr

Sara Saberi
Worcester Polytechnic Institute
Email: ssaberi@wpi.edu

ABSTRACT

This study investigates the potential of additive manufacturing (AM) technology in enhancing supply chain resilience (SCR) during periods of disruption. The impact of adopting AM on SCR is influenced by various factors, including institutional response, technological capabilities, manufacturing capabilities, and supplier relationships. The relationships between the variables were examined using structural equation modeling. The findings of the study confirm that AM technology has the ability to improve SCR, challenging the traditional perspective that institutional experience is the primary driver of preparedness. This research offers practical implications for managers seeking to enhance their organization's resilience within a dynamic business environment.

KEYWORDS: Supply Chain Resilience, Additive Manufacturing, Structural Equation Modelling

INTRODUCTION

In recent years, the world has witnessed several natural disasters, pandemics, and geopolitical tensions that have disrupted global supply chains. These disruptions have exposed the need for resilient supply chains that can withstand unexpected shocks and quickly adapt to changing circumstances. For instance, the 2011 earthquake and tsunami in Japan severely impacted the automotive industry because many of its suppliers were located in the affected areas. Similarly, the COVID-19 pandemic has disrupted supply chains globally, leading to shortages of critical medical supplies and equipment. Thus, SCR has become an essential consideration for businesses today. To mitigate the impact of disasters on supply chains, businesses need to adopt resilience strategies. These strategies involve building redundant supply chains, diversifying suppliers, and developing contingency plans that can be quickly implemented in case of disruptions (Gunasekaran et al., 2015).

There are several factors that contribute to the resilience of a supply chain, but some stand out more prominently than others. These include institutional response (IR) (Wu et al., 2023), technology capability (TC) (Naghshineh and Carvalho, 2022), manufacturing capability (MC) (Linkov et al., 2020), and supplier relations (SR) (Faruquee et al., 2021), all of which have a significant impact on the resilience of the supply chain. However, in addition to these factors, many Industry 4.0 technologies have emerged that can support or mediate these variables with the aim of enhancing SCR (Qader et al., 2020). It's worth noting that these technologies have the potential to bring about radical changes and affect existing operations management. Despite this, they are increasingly being adopted by businesses looking to improve their SCR and adapt to the ever-changing landscape of the industry.

Industry 4.0 technologies, such as the Internet of Things (IoT), big data analytics, and artificial intelligence (AI), offer new opportunities for building resilient supply chains. IoT sensors can provide real-time data on inventory levels, shipping conditions, and other factors that can impact supply chain performance. Big data analytics can help businesses identify potential disruptions and develop proactive strategies to mitigate them. AI can be used to optimize supply chain operations and improve decision-making (Spieske et al., 2021). The other Industry 4.0 technology AM is a manufacturing process that involves building products layer by layer from digital designs. This technology has several advantages for SCR. First, it enables on-demand production, which reduces the need for large inventories and can help businesses respond quickly to changing demand. Second, it allows for localized production, which reduces transportation costs and lead times. Third, it enables rapid prototyping, which can help businesses develop new products and designs quickly (Belhadi et al., 2022).

When it relates to SCR, it is well known that AM technology provides numerous benefits such as including tracing customer specifications and providing flexibility, letting re-designing and instantaneous manufacturing, nearness to the consumer and quick delivery times, adaptability at the spot of maintenance and service, less exposure to disruptions in supply chain processes, increasing service level with diverse product types, low inventory, and transportation coherence (Kunovjanek et al., 2022). In this regard, the relevant study has two key research questions, which are as follows:

RQ1: What is the impact of AM technology on SCR?

RQ2: What role does AM play in mediating between SCR and latent variables such as institutional response, manufacturing and technological capabilities and supplier relations?

AM technology can contribute significantly to enhancing the resilience of supply chains, and the research questions aim to explore the ways in which companies can leverage these benefits.

LITERATURE REVIEW

The scholars found that utilizing theoretical models could aid in elucidating the parties involved in adopting and executing technologies, as well as other essential factors, in order to discern and clarify the part played by governments, technologies, networks, stakeholders, and other establishments in implementing new technologies and ensuring the resilience of the supply chain (Pettit et al., 2010; Straub, 2009). In this section, we will explore the theories and associated presumptions that serve as the groundwork for our investigation while preserving the coherence of the meaning.

The stakeholder theory is a well-known management theory that emphasizes the importance of taking into account the impact of a company's activities on all stakeholders, not just its shareholders. These stakeholders can include employees, clients, suppliers, society, and the environment. To ensure supply chain resiliency, it is essential to consider the interests of all stakeholders in the supply chain, including governments and affiliated organizations (Qazi, et al., 2022).

Institutional theory suggests that external factors such as societal norms, values, and regulations can influence organizations. In the case of a supply chain disruption, government and other organizations can provide support that aligns with existing social norms and values. This support may include financial assistance, information exchange, regulatory flexibility, and coordination of response efforts. These actions are manifestations of societal standards such as organization, flexibility, cooperation, and teamwork. By providing this support, governments and other organizations can contribute to a more robust supply chain, reduce the impact of the disruption, and build trust with affected firms. Thus, it is critical to consider institutional reactions and supplier connections when analyzing the variables that impact the adoption and implementation of emerging technologies (Zsidisin et al., 2005).

Resource-based theory (RBV) is a management theory that suggests that a firm's resource and capabilities may provide it with a competitive advantage over the long term. Manufacturing capacity is one such capability that can be examined using the RBV framework, and it may provide a company with a competitive advantage. According to RBV, a company's resources can be classified into two categories: tangible and intangible. Tangible resources include things like machinery, equipment, and inventory, while intangible resources include knowledge, skills, and reputation. Manufacturing capacity includes both tangible and intangible resources such as inventory management procedures and machinery and equipment (Brandon-Jones et al., 2014).

Business network theory, which focuses on the relations between a company and its suppliers within a network of interdependent enterprises, also lends credence to the study's organizational framework. These relations are important because they may have a big impact on a company's access to resources, expertise, and information, as well as its overall competitiveness. According to business network theory, a company's supplier connections' strength and quality have an impact on how resilient the company is. A company may obtain essential resources and raw materials, simplify its supply chain, and save costs via improved coordination and communication by developing excellent supplier relationships. Moreover, a company may benefit from strategic supplier connections by gaining information and experience that it can use to boost product quality, innovation, and flexibility to market changes. Supplier relations are dynamic and prone to change over time, according to business network theory. Businesses need to be able to adjust to changes in their supplier relationships if they want to preserve resilience. Since they significantly affect a company's access to resources, expertise, and information as well as its competitiveness within a network of interdependent enterprises, supplier connections are a crucial component of business network theory. Businesses may increase their chances of success in the market by strategically managing their supplier relationships and adapting to changes over time (Mari et al., 2015).

To obtain valuable insights into the role of AM and other variables in enhancing SCR (Jain et al., 2018), we have employed a grounded theory research design in this study. The embedded theory approach allows us to create empirical theory and determine causal relationships between phenomena by analyzing qualitative data from the literature (Zsidisin et al., 2016). By analyzing the theoretical perspectives of resource-based theory, stakeholder theory, institutional theory, and business network theory, we have formulated hypotheses that address our research questions. Additionally, we have reviewed existing research on institutional interventions in disruption events, technology and production capabilities, supplier relationships, AM, and SCR to gain a comprehensive understanding of the research phenomenon. The combination of these analyses provides valuable insights into the role of AM and other variables in enhancing SCR.

THEORETICAL DEVELOPMENT/MODEL

SCR is the capacity of a supply chain to sustain operations and continue providing goods or services to clients in the face of unanticipated interruptions (Qader et al., 2022). In other terms, it refers to a supply chain's ability to quickly recover from or adjust to a disruption or shock. A supply chain that is resilient can recognize possible hazards and take proactive steps to reduce them. It entails having backup plans in place, dependable supplier connections, a diverse supplier pool, and efficient communication routes. In addition, a resilient supply chain is able to recognize and react to interruptions like natural disasters, cyberattacks, and geopolitical events promptly, reducing the effect on the whole supply chain and the end user. SCR has grown in importance as a means of preserving a competitive edge in the fast-paced, unpredictably changing business world of today (Tukamuhabwa et al., 2015).

In today's highly complex and uncertain business environment, SCR has become an essential concept for firms to survive and thrive (Gunasekaran et al., 2015). SCR has become an essential concept for firms to mitigate the negative effects of disruptions and maintain their competitive advantage. Institutional theory, stakeholder theory, and network theory provide valuable insights into the factors that influence a firm's SCR. Institutional theory suggests that firms are influenced by their institutional environments and must conform to institutional norms and values to gain legitimacy and support from stakeholders. This perspective suggests that firms with strong institutional legitimacy are more likely to have greater SCR, as they are better able to access resources and support from stakeholders (Zsidisin et al., 2005). Stakeholder theory posits that firms are interdependent with their stakeholders and must consider the needs and expectations of these stakeholders to maintain their social license to operate. The theory suggests that firms with strong stakeholder relationships are more likely to have greater SCR, as they can access resources and support from stakeholders during disruptions (Qazi et al., 2022). From the perspective of SCR, network theory suggests that supply chains are embedded in interdependent networks of relationships with other firms and stakeholders, and that these relationships can have a significant impact on SCR. According to this theory, a supply chain that is well-connected with its suppliers, customers, and other stakeholders is more likely to have greater resilience, as it can leverage these relationships to access resources and support during disruptions. Additionally, network theory emphasizes the importance of understanding the structure of a supply chain network, as it can help identify potential vulnerabilities and inform strategies for enhancing resilience. Overall, network theory provides a framework for understanding the complex interdependencies within a supply chain network and the role that these relationships play in shaping a supply chain's resilience. (Mari et al., 2015).

AM is a process of creating three-dimensional objects by adding layers of material one by one. This is in contrast to traditional manufacturing methods, which often involve removing material from a larger piece to create a desired shape. AM is also known as 3D printing, and it has been used in a variety of industries for prototyping, product design, and production of end-use parts. With AM, complex geometries can be created quickly and efficiently, which allows for greater flexibility and customization in manufacturing. The technology has the potential to significantly impact supply chain processes by reducing lead times, transportation costs, and material waste. Recent literature on AM has focused on its impact on supply chain management, including sustainability, spare parts supply chain, logistics, and supply chain design. Studies have highlighted the advantages of AM, including speed, flexibility, agility, and cost savings, but also noted challenges such as high equipment cost, processing time, and limited production capacity compared to traditional supply chains. Some researchers have also investigated the impact of AM on SCR, with findings suggesting that it can increase agility and responsiveness but also increase vulnerabilities. Overall, AM has the potential to improve supply chain processes and increase resilience, but careful consideration of its advantages and limitations is necessary (Verboeket et al., 2019).

In this respect, AM technology may have potentially positive effects on SCR (Naghshineh and Carvalho, 2022). Firstly, it allows for on-demand manufacturing, which can reduce lead times, inventory costs, and the need for stockpiling. This means that in times of supply chain disruptions, such as natural disasters or geopolitical events, companies can quickly adapt to changes in demand or supply by producing necessary parts or products in-house. Secondly, AM can enable distributed manufacturing, where parts or products can be produced closer to the end-users, reducing the reliance on centralized manufacturing and transportation networks. This can help mitigate risks associated with disruptions in transportation and logistics, as well as reduce carbon emissions associated with shipping products across long distances. Additionally, AM can facilitate design flexibility, as it allows for rapid prototyping and the ability to easily modify designs based on changing market or consumer needs. This can help companies quickly adapt to changing market conditions and maintain a competitive advantage. Thus, AM can play an important role in enhancing the resilience of supply chains by providing flexibility, reducing lead times, and enabling distributed manufacturing (Verboeket et al., 2019).

Conventional latent variables effect on SCR

Institutional Response: Institutional response is crucial for SCR and involves proactive planning, risk management, and clear communication channels. Organizations can have contingency plans to manage disruptions, while governments can contribute by implementing laws and regulations that enforce emergency response plans and frequent risk assessments. Regulatory organizations can also enforce laws and standards that focus on risk management, product safety, and quality control (Dubey et al., 2020; Scholten et al., 2019). Institutions are considered a latent variable in this study due to their significant impact on SCR and technological integrations. Additive manufacturing can also be considered within this context. Therefore, the impact of institutional response on AM integration can be considered from various perspectives. Consequently, we propose the following hypotheses:

H1a. Institutional response to disruptions in the supply chain has a positive effect on a firm's AM capabilities.

H1b. A firm's capability to improve SCR is positively influenced by institutional response to disruptions in the supply chain.

H1c. The mediating effect of AM technology capabilities positively influences the interaction between institutional response to disruptions in the supply chain and SCR.

Technology Capability: Technology capability is another important component of SCR. Organizations can leverage technology to improve visibility and transparency throughout their supply chains, which helps them identify potential risks and respond more quickly to disruptions. For example, a company might use sensors and data analytics to monitor their inventory levels and identify potential shortages or bottlenecks in their supply chain. Technologies of this kind can be evaluated in a broad range, particularly in the context of the Industry 4.0 ecosystem.

It is vital for businesses to be able to embrace and integrate new technology into their supply chain and manufacturing processes (Weigelt, 2009). Via technological maturity models, many criteria are utilized to analyze firms in this context (Kyriakidou et al., 2013). Technical capabilities may impact the integration process of new technologies like AM, as well as SCR, since enterprises without particular technological competencies may find it difficult to embrace AM technologies. As a result, the technical skills of businesses might have an influence on the application of AM technologies.

H2a. A company's technological capability to respond to supply chain disruption events has a substantial impact on its AM capabilities.

H2b. Technological capability of firms influences SCR significantly.

H2c. AM mediates the connection between technical competence and supply chain resiliency considerably.

Manufacturing Capability: Manufacturing capability is also critical for SCR. Organizations need to be able to quickly adjust their production schedules and capacity in response to disruptions. This requires flexible manufacturing processes and the ability to quickly retool production lines (Naghshineh and Carvalho, 2022). For example, a company might have the capability to switch production to different products or adjust their production schedules to accommodate unexpected changes in demand or supply. Thus, manufacturing capabilities of a company can have a significant impact on its SCR. On the other hand, companies with inferior manufacturing skills may struggle to immediately transfer output or boost production levels in response to

disturbances. Adopting AM technologies may be beneficial for firms with inadequate manufacturing capability features. AM technology can reduce a company's reliance on third-party suppliers and potential interruptions in its supply chain by enabling it to build components on demand and in-house, which increases flexibility and agility in reacting to changes in demand (Delic and Eyers, 2020; Bogers et al., 2016). In this context, we propose the following hypotheses.
H3a. Manufacturing capabilities influences SCR significantly.
H3b. The capacity of a company to respond to supply chain disruption events has a substantial impact on its AM capabilities.
H3c. AM mediates the link between manufacturing capabilities and SCR considerably.
Supplier Relations: Finally, supplier relations are a crucial component of SCR. Organizations need to work closely with their suppliers to ensure that they have the resources and capabilities to respond to disruptions (Kumar and Rahman, 2015). This includes developing strong partnerships with suppliers, sharing information about potential risks, and collaborating on contingency plans. Therefore, maintaining positive relationships with suppliers is crucial for the delivery of high-quality products and services at cost-effective prices (Patrucco et al., 2019). This partnership leads to higher product quality, lower prices, more efficiency, and more innovation. Good supplier connections also help to secure the availability of critical materials and equipment, which can benefit SCR (Kumar and Rahman, 2015). Furthermore, the communication, trust, and collaboration with suppliers can have a direct or mediating impact on supplier resilience. AM technologies can act as a trigger for investment and integration in supplier relationships, as they increase the direct impact of supplier relations on SCR. In any disaster scenario, the distributed nature of AM technologies can increase suppliers' collaborative competence and avoid long-term supply shortages, leading to an increase in supplier resilience. This study analyzes the mediating role of supplier relations that affect SCR. Because of this, we will take into account the following hypotheses:
H4a. Supplier relationships have a substantial influence on SCR.
H4b. A company's supplier relations' reaction to events that disrupt the supply chain has a major influence on that company's AM capabilities.
H4c. AM strongly mediates the link between supplier relationships and SCR.

METHODOLOGY

Research design

The quantitative methodology was utilized in this research to assess the model presented in Figure 1. This involved creating a survey instrument and using SEM to analyze the anticipated relationships. A questionnaire-based survey approach was selected since it allows researchers to gather and examine connections between several variables on a large sample size, enhancing the generalizability of the findings (Straub et al., 2004). Churchill (1979) notes that when designing reliable scientific measurements, the first step is to define the domain of the construct, which necessitates a literature survey.
As an initial step, we developed the model and elements outlined in Figure 1, and then relied on the available body of literature to devise the survey tool (Moore and Benbasat, 1991). The questionnaire employs a seven-point Likert scale to gather responses across various dimensions. Multiple aspects are assessed through this survey approach, allowing for a comprehensive evaluation.

Data collection

To gather necessary information for our research goals, we developed a survey questionnaire using expert opinions and relevant studies. The questionnaire was drafted in Turkish and later translated to English, then reviewed by two professors. Feedback was received on the appropriateness and clarity of the survey, and corrections were made with the help of manufacturing managers. We then conducted interviews with professionals in AM and carried out a pilot study with 27 industry workers to refine the questionnaire. Then, we collected data from 217 (348 professionals reached) management professionals in Turkey through an internet survey platform, resulting in a response rate of 62%. After removing incomplete responses, we evaluated the data set for analysis to apply structural equational modeling.

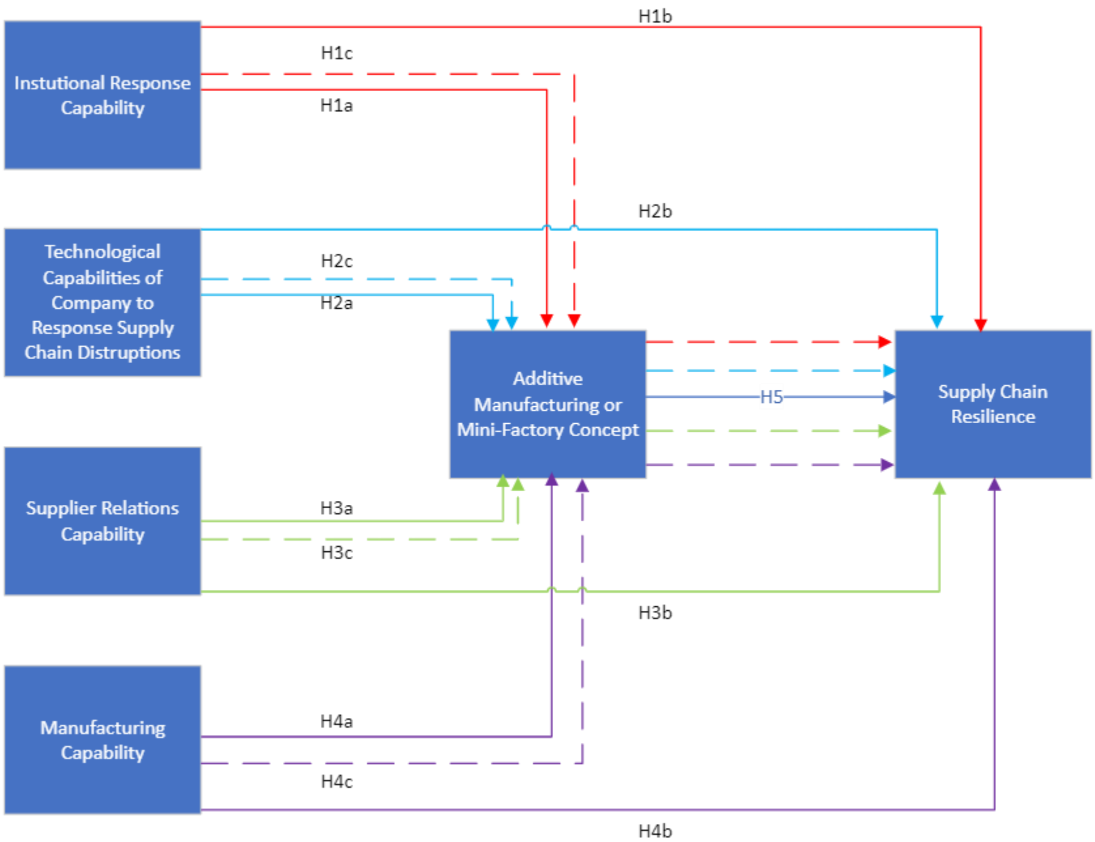
Model analysis

The analysis of the data obtained through the survey consists of two stages: validity and reliability, and structural model. To create the structural model, the model must first be validated. Thus, to determine the viability of the concept being tested, we conducted an exploratory factor analysis (EFA) using principal component analysis and promax rotation approach. All items on the questionnaire were input simultaneously, and as hypothesized, the pattern matrix revealed six distinct constructs. None of the items were deleted, as all had factor loadings above 0.40. Additionally, all six constructs had eigenvalues above the acceptable threshold value 1. A confirmatory factor analysis (CFA) was done using Amos 22 to further establish the concept's applicability. There was no requirement to eliminate any more components to enhance the model fit indices since they were already over the permitted threshold values. The final model fit indices were adequate and offered a decent fit for the data. They were based on first-order constructs. After ensuring the reliability and validity of the model through the analysis, we proceeded to the second stage.

In the second phase (structural model), to test the hypothesized relationships depicted in Figure 1, we utilized SEM, which allows for simultaneous assessment of both direct and indirect effects. Furthermore, the bootstrapping approach was utilized, as it enhances the accuracy of the analysis. To address Common Method Bias (CMB), bias corrected imputed component scores produced previously were employed. Before including the moderating effect on the relationship in the model, the direct impact of the independent variables; institutional response (IR), technological capability (TC), manufacturing capability (MC), and supplier relations (SR) on SCR was explored. Positive and significant standardized regression weights were found for the relationships between independent variables (IR, TC, MC, and SR) and the dependent variable (SCR). These findings support the hypotheses H1b, H2b, H3b, and H4b.

Moreover, the research claimed that AM mediates the associations between the dependent variable (SCR) and many independent variables (IR, TC, MC, and SR). The study used a mediation analysis with the bootstrapping technique to test the proposed mediating role of AM in the relationship between IR, TC, MC, SR, and SCR. The results showed that IR, TC, and MC were positively associated with AM, supporting H1a, H2a, and H3a hypotheses, respectively. However, SR was not found to have a significant direct effect on AM, and therefore H4a was not supported. Also, AM was found to have a positive and significant direct effect on SCR, supporting H5a. The bootstrapping results revealed that AM partially mediated the relationship between TC and SCR, while fully mediating the relationship between MC and SCR. Therefore, H2c and H3c were supported, but H1c and H4c were not supported.

Figure 1: Structural Equation Model



DISCUSSIONS

The study empirically tests the impact of manufacturing capability, technological capability, institutional response, and supplier relations on SCR and highlights the crucial role of AM in enhancing SCR. The study provides initial evidence on the mediating effects of AM in the relationship between these variables and business performance. From a managerial perspective, the study suggests that companies need to improve their existing capabilities, invest in new technologies such as AM, establish crisis management teams, establish communication protocols among supply chain partners, and establish long-term relationships with suppliers to enhance SCR. However, the implementation of AM requires careful planning and execution, including a detailed analysis of relevant business processes, evaluation of investment costs, training of personnel, implementation of strict safety and quality control measures, and integration with supply chain management.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

Scholars have been working to identify different organizational, technical, financial, and managerial capabilities that might assist organizations in mitigating the detrimental effect of supply chain disruptions in recent years. This study adds to the expanding body of research in this area by underlining the critical role of AM technological capabilities in helping organizations to improve their capacity to deal with supply chain risks. The study's results show that developing AM skills improves a company's ability to use its in-house manufacturing capability to minimize future supply chain interruptions. Moreover, the findings highlight the importance of AM skills in enhancing the impacts of institutional preparedness, technical competence, and production capability on a firm's ability to handle supply chain disruptions. Moreover, we argue that AM technology directly helps to SCR. Based on our results, we propose that businesses prioritize the integration of AM technology inside their companies in order to successfully increase their supply chain risk mitigation capabilities.

While this study contributes to our understanding of SCR and the role of AM and the other latent variables, it is important to acknowledge its limitations. In this study, we focus primarily on Turkish enterprises, and so may not fully reflect the viewpoints of managers from Asia, the United States, or Europe, who may have different ideas on the relevance of AM skills in improving SCR. Second, since the research includes a wide variety of industries, we are unable to give industry-specific insights on how AM might be used to enhance risk resilience. Despite these limitations, the research provides significant information and serves as a starting point for further understanding how AM may help organizations improve their SCR. Additionally, AM technologies have the potential to be widely used in the service industry, such as healthcare. As a result, relevant research might be conducted, taking into consideration various points of view.

REFERENCES

Belhadi, A., Kamble, S. S., Venkatesh, M., Jabbour, C. J. C., & Benkhati, I. (2022). Building supply chain resilience and efficiency through additive manufacturing: An ambidextrous perspective on the dynamic capability view. International Journal of Production Economics, 249, 108516.

Bogers, M., Hadar, R., & Bilberg, A. (2016). Additive manufacturing for consumer-centric business models: Implications for supply chains in consumer goods manufacturing. Technological forecasting and social change, 102, 225-239.

Brandon-Jones, E., Squire, B., Autry, C. W.,Petersen, K. J. (2014). A contingent resource-based perspective of supply chain resilience and robustness. Journal of Supply Chain Management, 50(3), 55-73.

Churchill Jr, G. A. (1979). A paradigm for developing better measures of marketing constructs. Journal of marketing research, 16(1), 64-73.

Delic, M., & Eyers, D. R. (2020). The effect of additive manufacturing adoption on supply chain flexibility and performance: An empirical analysis from the automotive industry. International Journal of Production Economics, 228, 107689.

Dubey, R., Gunasekaran, A., Bryde, D. J., Dwivedi, Y. K., & Papadopoulos, T. (2020). Blockchain technology for enhancing swift-trust, collaboration and resilience within a humanitarian supply chain setting. International journal of Production research, 58(11), 3381-3398.

Faruquee, M., Paulraj, A., & Irawan, C. A. (2021). Strategic supplier relationships and supply chain resilience: Is digital transformation that precludes trust beneficial?. International Journal of Operations & Production Management, 41(7), 1192-1219.

Gunasekaran, A., Subramanian, N., & Rahman, S. (2015). Supply chain resilience: role of complexities and strategies. International Journal of Production Research, 53(22), 6809-6819.

Jain, V., Kumar, S., Soni, U., & Chandra, C. (2017). Supply chain resilience: model development and empirical analysis. International Journal of Production Research, 55(22), 6779-6800.

Kumar, D., & Rahman, Z. (2015). Sustainability adoption through buyer supplier relationship across supply chain: A literature review and conceptual framework. International strategic management review, 3(1-2), 110-127.

Kunovjanek, M., Knofius, N., & Reiner, G. (2022). Additive manufacturing and supply chains–a systematic review. Production Planning & Control, 33(13), 1231-1251.

Kyriakidou, V., Michalakelis, C., & Sphicopoulos, T. (2013). Assessment of information and communications technology maturity level. Telecommunications Policy, 37(1), 48-62.

Linkov, I., Carluccio, S., Pritchard, O., Ní Bhreasail, Á., Galaitsi, S., Sarkis, J., & Keisler, J. M. (2020). The case for value chain resilience. Management Research Review, 43(12).

Mari, S. I., Lee, Y. H., & Memon, M. S. (2015). Complex network theory-based approach for designing resilient supply chain networks. International Journal of Logistics Systems and Management, 21(3), 365-384.

Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. Information systems research, 2(3), 192-222.

Naghshineh, B., & Carvalho, H. (2022). The implications of additive manufacturing technology adoption for supply chain resilience: A systematic search and review. International Journal of Production Economics, 247, 108387.

Patrucco, A. S., Luzzini, D., Moretto, A., & Ronchi, S. (2019). Attraction in buyer–supplier relationships: Improving supply network performance through purchasing recognition and proficient collaboration initiatives. Business Process Management Journal, 25(2), 347-367.

Pettit, T. J., Fiksel, J., & Croxton, K. L. (2010). Ensuring supply chain resilience: development of a conceptual framework. Journal of business logistics, 31(1), 1-21.

Scholten, K., Sharkey Scott, P., & Fynes, B. (2019). Building routines for non-routine events: supply chain resilience learning mechanisms and their antecedents. Supply Chain Management: An International Journal, 24(3), 430-442.

Spieske, A., & Birkel, H. (2021). Improving supply chain resilience through industry 4.0: A systematic literature review under the impressions of the COVID-19 pandemic. Computers & Industrial Engineering, 158, 107452.

Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. Communications of the Association for Information systems, 13(1), 24.

Straub, E. T. (2009). Understanding technology adoption: Theory and future directions for informal learning. Review of educational research, 79(2), 625-649.

Tukamuhabwa, B. R., Stevenson, M., Busby, J., & Zorzini, M. (2015). Supply chain resilience: definition, review and theoretical foundations for further study. International journal of production research, 53(18), 5592-5623.

Qader, G., Junaid, M., Abbas, Q., & Mubarik, M. S. (2022). Industry 4.0 enables supply chain resilience and supply chain performance. Technological Forecasting and Social Change, 185, 122026.

Qazi, A. A., Appolloni, A., & Shaikh, A. R. (2022). Does the stakeholder's relationship affect supply chain resilience and organizational performance? Empirical evidence from the supply chain community of Pakistan. International Journal of Emerging Markets, (ahead-of-print).,

Verboeket, V., & Krikke, H. (2019). The disruptive impact of additive manufacturing on supply chains: A literature study, conceptual framework and research agenda. Computers in Industry, 111, 91-107.

Weigelt, C. (2009). The impact of outsourcing new technologies on integrative capabilities and performance. Strategic Management Journal, 30(6), 595-616.

Wu, Q., Zhu, J., & Cheng, Y. (2023). The effect of cross-organizational governance on supply chain resilience: A mediating and moderating model. Journal of Purchasing and Supply Management, 100817.

Zsidisin, G. A., Melnyk, S. A., & Ragatz, G. L. (2005). An institutional theory perspective of business continuity planning for purchasing and supply management. International journal of production research, 43(16), 3401-3420.

Zsidisin, G.A., Petkova, B., Saunders, L.W. and Bisseling, M. (2016), “Identifying and managing supply quality risk”, International Journal of Logistics Management, Vol. 27 No. 3, pp. 908-930.

DECISION SCIENCES INSTITUTE

An Adaptive Procedure for Intermittent Forecasting: A Simulation Comparison of Methods

Matthew Lindsey
Stephen F. Austin State University
Email: lindseymd@sfasu.edu

Robert Pavur
The University of North Texas
Email: pavur@unt.edu

ABSTRACT

Items with intermittent demand are often subject to changing demand and even to obsolescence. Croston's Method and its variations, have been demonstrated to not provide particularly accurate forecasts in cases of decreasing demand that leads to obsolescence. Generally, when the distribution of demand changes from a high demand to a low demand state and perhaps going back to a high demand state, forecasting methods must be adaptive. The literature is rich with variations of intermittent forecasting methods, but very limited in studying conditions with changing demand distributions. An adaptive approach to Croston's method is shown to be an improvement and a simulation experiment reveals the conditions under which it outperforms the traditional Croston method as well as SES.

KEYWORDS: Forecasting, Fires, Croston's method, simulation, Single Exponential Smoothing

INTRODUCTION

A key characteristic of intermittent demand is to have a large number of time periods with no demand. Examples of these items might be spare parts, rare or exclusive items, or items that are near obsolescence. Inventory system generally attempt to minimize costs for these items, then a suitable forecast is need for future average demand. Often, products only have positive demand in a few periods and then zero demand. Generally, more periods of zero demand, will increase the forecast error. Syntetos, Boylan and Croston (2005) used demand classification to help identify the optimal forecasting methodology. Special cases like obsolescence, shifting demand or the presence of trends can cloud the advantages of even optimal methodologies.

SES is a dependable technique to forecast when demand rates are stable. Croston's (1972) variation of SES for intermittent demand cases is the principal procedures for forecasting slow-moving items. Advantages are identified by Willemain, Smart, Shockor, and DeSautels (1994). Johnston and Boylan (1996), Syntetos and Boylan (2005) demonstrate its dominance. Studies of seasonality, promotions, irregular events, trends and correlated demands on the accuracy of forecasts have been conducted (Altay, Litteral and Rudisill, 2012; Lindsey and Pavur, 2008).

LITERATURE REVIEW

Slow-Moving Inventory

According to the literature, selecting the appropriate forecasting and stock-control methodology requires that the practitioner select the correct demand pattern (Syntetos, Boylan, and Croston, 2005; Boylan, Syntetos and Karakostas, 2006). Only by suitably categorizing demand can the best available forecasting method be selected (Syntetos, et al 2005, Rožanec, Fortuna, and Mladenić, 2022). Syntetos, Boylan and Croston (2005) segmented groups based on the time between demands and the coefficient of variation of the demand amount squared, but no consensus exists that it is the best methodology for sorting demand.

Assumed Distributions

The assumed primary distribution for time between positive demands is the geometric distribution in the literature (Croston, 1972, Segerstedt, 1994, Willemain, et al, 1994, Bagchi, havya and Ord, 1983, Leven and Segerstedt, 2004). Various articles have investigated demand size and demand size distribution issues and generally assume a normal distribution (Croston, 1972, Segerstedt, 1994, Willemain, et al, 1994, Bagchi, havya and Ord, 1983) however the lognormal and Erlang have been used as well (Willemain, et al, 1994 and Leven and Segerstedt, 2004). The authors have investigated scenarios that call for changing forecasting methodologies as the demand changes. The assumption of geometric fails when the demand type is changing but it is not clear when the rate decreases, as in obsolescence of the impact.

Obsolete Inventory

The effect of obsolescence on forecasting intermittent demand has been examined in a few cases. A technique was introduced by Teunter, Syntetos and Babai (2011) for inventory obsolescence with intermittent demand. Babai, Syntetos and Teunter (2014) looked at the effects of smoothing constants, when obsolescence is expected. Babai, Dallery, Boubaker and Kalai (2019) presented a method as well when obsolescence is expected but called for more research due to conflicting results but does not consider the impact of intermittency, lumpiness and non-stationarity. Pince and Dekker (2011) investigated a continuous review inventory system with expected obsolescence and Van Jaarsveld and Dekker (2011) proposed methodology to determine obsolescence risk to improve inventory systems. Sanguri, Patra, and Punia, (2023) investigated using temporal hierarchy, which uses the structural connection across levels of aggregation, when forecasting intermittent demand with the risk of obsolescence. The study shows the Croston techniques do not do very well when trends or seasonality is present, as expected.

Intermittent Demand Forecasting Method

The seminal work for forecasting intermittent demand is Croston (1972) who created two time series, one for demand and one for the time between demand. Traditionally, SES, a weighted-moving average technique used to forecast demand by using exponentially decreasing weights over time, is recommended for stationary time series. Willemain, et al (1994) should be consulted for a full explanation of Croston's method.

Considerable research has been done on Croston's method including Syntetos and Boylan(2005) who provided a bias correction for Croston's method known as the Syntetos and Boylan Approximation (SFA). The SBA expected estimate of demand is then:

$$\mathrm{E}(Y_t') = \mathrm{E}\left(\left(1 - \frac{\alpha}{2}\right)\frac{z_t'}{p_t'}\right) \approx \frac{\mu}{p} - \frac{\alpha}{2}\frac{\mu}{p^2} \tag{1}$$

Teunter et al. (2011) proposed a modification to be used when the possibility of obsolescence exists. TSB updates the probability of demand instead of the inter arrival time in every period and is always updated even if many periods of no demand occur. It is unbiased and has works well with obsolescence but does not consistently beat the SBA methodology (Babai et al 2014).

Babai, Dallery, Boubaker and Kalai (2019) proposed a modification to the Syntetos and Boylan Approximation that updates the demand size, the demand interval and the estimator in periods with a demand. However, as the number of periods between positive demand episodes increases the update becomes similar to the probability of demand estimate in the (TSB) Method.

Proposed Adaptive Method

A proposed method of incorporating an adaptive exponential smoothing method into Croston's procedure is presented. As seen in Croston's method, the exponential smoothing method is used only when demand has occurred. In this proposed method, the smoothing constant is allowed to change. This change will occur according to some proposed guidelines using time between demands. Let Q represents the time between the current demand and the last demand. The algorithm for allowing the smoothing constant to change in our study is the following.

1. When a demand has occurred, compute Diff_Q to be the difference between the time since the last demand and the previous demand. The difference is computed to be the absolute value of the difference between the last two times between demands.
2. Compute SM_Diff_Q to be the single exponential smoothed value of Diff Q using the same smoothing constant intended to be used in Croston's procedure. Label Prev_SM_Diff_Q the previous smoothed value of SM_Diff_Q and the time of a demand.
3. Compute the absolute value of Diff_Q Change the smoothing constant according to:
a. If abs(Diff_Q – Prev_SM_Diff_Q) ≤ 1.5, then smoothing constant does not change.
b. If abs(Diff_Q – Prev_SM_Diff_Q) > 1.5, then new smoothing constant is equal to .2.
c. If abs(Diff_Q – Prev_SM_Diff_Q) > 2.0, then new smoothing constant is equal to .3.
d. If abs(Diff_Q – Prev_SM_Diff_Q) > 3.0, then new smoothing constant is equal to .4.
e. If abs(Diff_Q – Prev_SM_Diff_Q) > 4.0, then new smoothing constant is equal to .5.

The concept supporting this algorithm is that a larger smoothing constant allows for the smoothing procedure to adjust faster to the changing distribution of the data. However, sometimes, the distribution is not changing but there are sporadic changes in the time between demand. In this case, the proposed algorithm for changing the smoothing constant will decrease the performance of the smoothing procedure. If the distribution of the time between demands is indeed changing, then the proposed algorithm will enhance the performance of the smoothing procedure. This concept is not new. Adaptive exponential smoothing has many cites in the literature. Ekern (1981) explained its benefits 40 years ago. Despite the intuitive appeal of this procedure, limited research regarding enhancing Croston's procedure with it is published.

Simulation Experiment

To assess the proposed adaptive method, a simulation study is conducted in which the distribution of the time between demands changes. To enhance the ease of describing the experiment, two states are created. One state is considered the fast demand state and the other is considered the slow demand state. The standard assumptions of Croston's method are assumed, except that the state of demand can change from fast to slow. The number of periods until a demand is assumed to follow a geometric distribution with a fixed probability. The demand is independent of the time till demand and is assumed to follow a normal distribution.

Two states are simulated – fast and slow. The fast demand state has a relatively high probability of a demand occurring of .95, .8, or .5 and the slow demand state has a relatively low probability of a demand occurring of .3. The fast state will have a demand that follows a normal distribution with a mean of 500 and standard deviation of 100. The slow state will have a demand that follows a normal distribution with a mean of 200 and a standard deviation of 20. The number of periods of a fast demand is 50 and it is followed by 100 periods of slow demand. The total time period for each set of data in which a forecasting method was evaluated was 5*(50+100) = 750 time periods in one time period of data. These data sets were replicated 500 times.

The smoothing constants are either .08, .05, and .01. Four forecasting methods are used in the simulation study. SES and Croston's method are included. The proposed adaptive method using Croston's procedure is included and is labeled "Croston D_SM". When "Croston D_SM" uses the SBA bias correction procedure, the method is labeled "Croston D_SM Bias_C." Table 1 shows RMSE accuracy in estimating the true mean demand during the fast or slow periods. The exception is the last three rows in Table 1 which omit the slow demand state.

Table 1. RMSE performance of four methods to model changing intermittent demand. Optimal procedures for each condition are in bold.

Smoothing constant	**Slow Demand Prob**	**Fast Demand Prob**	**Single Exponential Smoothing**	**Croston**	**Croston D_SM**	**Croston D_SM Bias_C**
0.08	0.3	0.95	**118.0066**	156.9908	119.5306	119.0143
0.08	0.3	0.8	**99.7604**	133.5428	103.2994	102.6988
0.08	0.3	0.5	**62.9336**	80.5512	71.5385	70.3874
0.05	0.3	0.95	142.0684	175.2600	**126.4650**	126.4667
0.05	0.3	0.8	118.0537	146.6261	**109.6647**	109.3775
0.05	0.3	0.5	**69.7705**	85.4821	76.4650	75.6724
0.01	0.3	0.95	198.1356	**194.5349**	200.0393	200.2329
0.01	0.3	0.8	162.5797	160.2244	**155.8886**	156.0427
0.01	0.3	0.5	91.3467	92.3775	83.6891	**83.6774**
0.08	No Slow	0.8	63.1049	**59.0075**	70.314394	72.412148
0.05	No Slow	0.8	56.4225	**53.7365**	73.921136	76.202944
0.01	No Slow	0.8	46.1057	**45.5283**	106.16637	106.94293

RESULTS

Table 1 illustrates that the proposed adaptive method is not always advantageous. The last three conditions were included to determine the detrimental effect of the proposed adaptive method when it is not needed. Croston's method with not modifications should perform the best and it does. However, the proposed adaptive procedure underperforms both Croston's unmodified method and single exponential smoothing. These last three rows confirm that Croston's method outperforms the exponential smoothing method even with a modestly high probability of demand occurrences equal to .8. For lower probabilities of demand occurrences, the benefit of Croston's procedure increases and is not provided in the table.

Table 1 clearly demonstrates that Single Exponential Smoothing excels when there are changing periods of fast and slow demand distributions and when the smoothing constant is .08. However, the Croston D_SM and Croston D_SM Bias C are both much closer to the Single Exponential Smoothing method then Croston's method. There is a very slight benefit to the bias correction procedure.

When the smoothing constant is .05, Croston D_SM appears to have substantial advantage, along with its bias correction version. However, when the probability of fast demand occurrences drops to .5, Single Exponential Smoothing again excels. When the smoothing constant is .01, Croston D_SM and Croston D_SM Bias_C prove to be beneficial.

CONCLUSIONS

While various modifications of Croston's method have emerged in the literature, no proposed procedures with distinct guidelines for improving Croston's procedure with respect to changing the smoothing constant are provided. Research with regard to obsolescence has been proposed and this research is clearly addressing a changing distribution for the demand and its occurrences. In the simulation, the states studied are reflective of a condition in which products often enter the market at a high price and then when the next version of the product or of a competitor's product emerges, the price of the older versions decrease and sell at a slower rate. This effect is clearly visible in the automotive market and in the cell phone market.

The results of the simulation study reveal that the dynamic nature of changing the smoothing constant for changing environments can have a substantial benefit when applied to Croston's procedure. Unfortunately, predicting the changing environments of markets can be tricky and subjective. No forecasting method for intermittent data has been shown to excel under all conditions. The condition of changing demand environments is indeed a challenging condition to impose on any forecasting procedure. Further simulations and studying the proposed adaptive method to Croston's procedure using real world is the next step for further research.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

An Efficiency-Based Clustering Analysis Approach to Emergency Logistics Network System Design

Jae-Dong Hong
Industrial Engineering
South Carolina State University
Orangeburg, SC 29117
Email: jhong@scsu.edu

ABSTRACT

The emergency logistics network system (*ELNS*) design has become a significant issue due to recent pandemic outbreaks and global natural/climate-related disasters. The *ELNS* design has recently required multi-objective models and has been evaluated by the data envelopment analysis (DEA)-based methods. This paper develops and proposes a new method, the efficiency-based clustering analysis (EBCA) method, to provide a better design framework. We demonstrate this design framework would help the decision-makers better evaluate the efficiency of various *ELNS* configurations than DEA-based methods and identify the efficient and robust ones among them. A case study compares the results between EBCA and DEA-based methods.

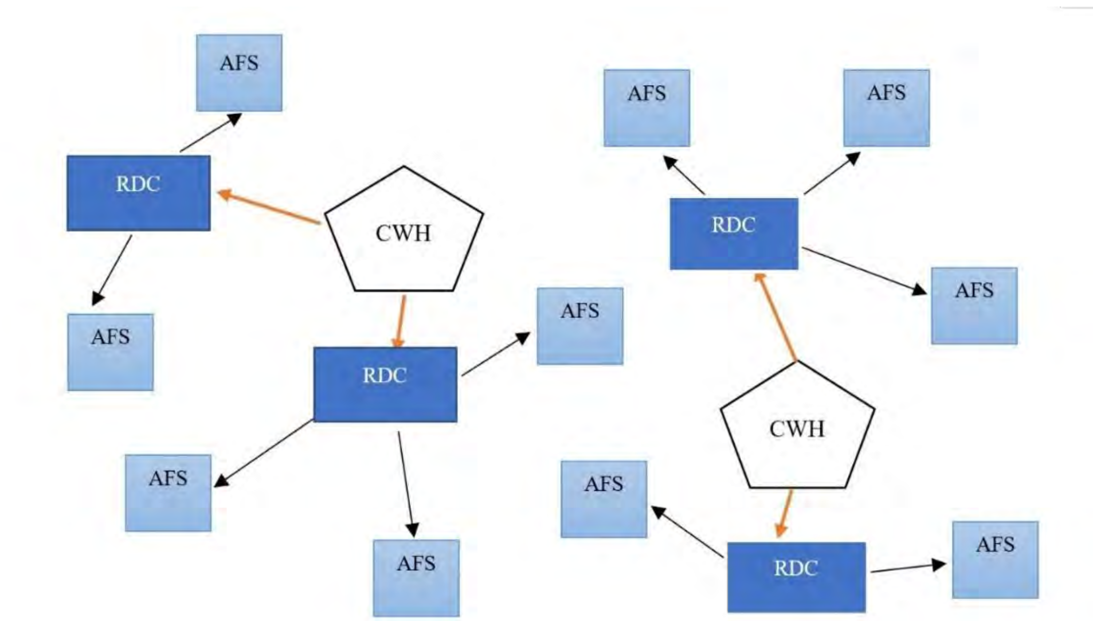
KEYWORDS**:** Emergency logistics network system, Data envelopment analysis, Multi-objective models, Efficiency-based clustering analysis

INTRODUCTION

The recent COVID-19 outbreak and severe weather-natural disasters make the design of the emergency logistics network system (*ELNS*) a critical strategic issue for the world in a pre-disaster scenario. The *ELNS* design problem deals with the location-allocation of emergency response facilities (ERFs). Creating and maintaining the efficient and effective *ELNS* is essential in providing vaccines or relief items such as food, first aid, drinking water, and daily commodities to alleviate people's suffering (Day et al., 2012; Boonmee et al., 2017; Shavaranim, 2019). This paper proposes and demonstrates how to design an efficient *ELNS* configuration under the risk of ERF disruptions. For this purpose, consistent and robust performance evaluation is essential. The *ELNS* considered in this paper is a two-echelon supply chain system with three distinctive ERFs, as shown in Figure 1. They are (i) Central Warehouses (CWHs), where humanitarian items are stored, (ii) intermediate response facilities termed Relief Distribution Centers (RDCs), where disaster-affected people can gain access to relief goods efficiently, and (iii) Affected Sites (AFSs) that would need humanitarian items. The main target of the strategic level is to fortify emergency preparedness and select the most cost/distance-efficient and suitable location of RDCs and CWHs among a set of candidate locations. In addition, another target is to establish the distribution of humanitarian supplies throughout the *ELNS* and to allocate AFSs to RDCs and RDCs to CWHs. However, traditional cost-based ERF location-allocation models implicitly assume that all ERFs will always provide service or are available and do not consider an associated risk of facility disruption. Due to natural disasters, pandemics, accidents, or strikes, all ERFs are open to disruptions. Such troubles would be worsened due to a lack of flexibility and interdependency, commonly presented in the general supply chain systems.

Evaluating various *ELNS* alternatives and identifying the most efficient schemes would be critical for logistics network planning. The multi-objective programming (MOP) models request decision-makers to assign different weights to the deviational variable of each objective. It would require decision-makers to reflect on the importance and desirability of deviations from the multiple goals. However, the actual efficiency of the *ELNS* generated by solving the model with a fixed weight is unknown. No standard procedure for assigning weight values is available to make sure that decision-makers will find the most desirable solution to an MOP problem. Ragsdale (2022) suggests that an iterative procedure should be followed, using a particular set of weights, concluding that we must repeat this procedure multiple times to find the most needed solution for decision-makers. Consequently, it is inevitable for decision-makers to use some of their subjective judgment. Evaluating various *ELNS* network schemes objectively, not subjectively, and selecting the most efficient alternatives would be essential for designing the *ELNS*. Hence, we should be able to propose a better procedure for assessing the efficiency of all options generated by the MOP model and select the most desirable one(s) without any subjective judgment.

Figure 1: Two-echelon emergency logistics network system



The data envelopment analysis (DEA) method was developed to evaluate the relative efficiency of decision-making units (DMUs) in the presence of multiple outputs to produce and multiple inputs to consume. The traditional DEA (T-DEA), proposed by Charnes et al. (1978), yields a single, comprehensive performance measure, which is called efficiency score (ES), for each DMU. The efficient ones would be identified with the ES of one (1.0). The T-DEA, with the principle of self-evaluation, permits each DMU to be rated with its most favorable weights. Thus, each DMU is allowed to disregard unfavorable outputs/inputs to maximize its self-efficiency under the T-DEA model. Consequently, the lack of discrimination is the most critical weakness of T-DEA. Many DEA-based models have been developed. The cross-efficiency (CE) DEA method is the most widely applied one to remedy this crucial deficiency of T-DEA.

The CE-DEA assesses DMUs with the main idea of peer evaluation rather than T-DEA's self-evaluation. Due to its enhanced discriminating power, many applications based on the CE evaluation have been proposed (see Sexton et al., 1986; Paryzad et al., 2018; Lee, 2019; Liu et al., 2019). As Doyle and Green (1994) note, the first issue of applying the CE-DEA model is the non-uniqueness of cross-efficiency scores (CESs) due to the often-present multiple optimal DEA weights. The second issue is that the CE method frequently ranks inefficient DMUs higher than fully efficient DMUs identified by T-DEA. In the meantime, Anderson and Peterson (1993) develop the idea of super-efficiency (SE) by excluding a DMU under evaluation from the reference set of the T-DEA model. However, the critical issue of using the SE-DEA model is that the adjacent DMUs determine a DMU's super-efficiency score (SES), so it would sometimes be unreasonable for DMUs to be evaluated and ranked by the SESs.

Clustering analysis (CA), a data mining technique, is an exploratory data analysis tool for solving grouping/classification problems. Clustering aims to identify meaningful segmentations or groupings of entities within a data set, focusing on distance-based analysis. In other words, clustering sorts entities into clusters or groups so that the degree of association is substantial among members of the same cluster and weak among members of different clusters (see Delen, 2021). One of many tools for calculating distance measures is the centroid-based technique. We modify the original clustering analysis regarding clustering entities and propose an **efficiency-based clustering analysis** (EBCA) method using the Euclidean distance. What distinguishes the proposed EBCA method from other clustering analysis methods is that the EBCA utilizes the minimum and maximum points to represent the clusters rather than the centroid of each cluster. The EBCA method divides all DMUs into efficient and inefficient clusters, as T-DEA does, and allows some DMUs to belong to both clusters. If DMU's inputs are close to the minimum point, it would belong to the efficient group regarding its inputs. Regarding outputs, the DMU would belong to the efficient group if they are close to the maximum point. Contrary to the principle of the DEA-based models, the EBCA method does not ignore or sacrifice any unfavorable inputs/outputs. The proposed method would eliminate the critical issues of DEA-based evaluation models and rate DMUs more impartially and consistently than the DEA-based methods.

As Big Data research becomes a vital area of operations analytics, DEA is evolving into data-enabled analytics and a data-oriented data science tool for various operational analytics. Various researchers show that Big Data can be applied to improve company productivity or efficiency and will be essential for enterprises to grow and achieve a competitive advantage. When researchers attempt to integrate DEA into Big Data, the serious barrier between Big Data and DEA is the excessive running time for solving a large set of DMUs in the Big Data context. The proposed EBCA process takes advantage of the simple computation, not requiring any optimization software that all DEA-based models need. Consequently, the proposed method would be a more appropriate tool for solving a large set of DMUs in the Big Data context. Therefore, the EBCA method could help practitioners and researchers produce a more refined and consistent efficiency evaluation than the DEA-based methods. Besides, the proposed procedure would quickly provide a benchmarking framework for DMUs to improve efficiency.

The significant contribution of this study is to propose an innovative framework consisting of how to formulate the design problem as an MOP model and evaluate and identify the robust *ELNS* schemes for efficient disaster/emergency relief operations. The EBCA method can allow us to assess various *ELNS* configurations objectively, not subjectively, and select the most efficient alternatives, which would be essential for planning the *ELNS* schemes. This paper will first attempt to apply the proposed EBCA method for designing *ELNS* and compare the results with those generated by CE-DEA, the most popular DEA-based method.

We apply the ELNS model that Hong and Jeong (2019) consider for comparison purposes. For planning more balanced ELNS schemes, weighted goal programming (WGP) is applied to generate various network alternatives. Theoretically, the weight assigned to each objective should be a continuous variable between 0 and 1. Instead of continuous variables, we limit each weight to discrete values, changes between 0 and 1 with an increment of 0.1 for a practical purpose. Solving the above WGP model for a given set of weights generates one ELNS scheme with a group of optimal four-performance measures. There will be various ELNS configurations depending on the values of the set of weights, $\boldsymbol{\alpha}$. To evaluate these generated ELNS schemes, we apply CE-DEA and the proposed EBCA method and compare the results of these two methods.

DATA ENVELOPMENT ANALYSIS METHODS

We apply the ELNS model that Hong and Jeong (2019) consider. Note that each DMU with two inputs, *EDC* (Expected Demands Covered) and *CDE* (Covered Demands in case of Emergency), and two outputs, *TLC* (Total Logistics Cost)and *MCD* (Maximum Coverage Distance), represents an *ELNS* scheme yielded by solving the WGP model for a fixed value of each weight. Letting E_t represent the efficiency score (ES) for DMU_t, we formulate the following linear programming model of T-DEA for DMU_t with two outputs and two inputs as follows:

$$Max\, E_t = u_{1t}EDC_t + u_{2t}CDE_t, \tag{1}$$

subject to

$$v_{1t}TLC_t + v_{2t}MCD_t = 1, \tag{2}$$
$$(u_{1t}EDC_d + u_{2t}CDE_d) - (v_{1t}TLC_d + v_{2t}MCD_d) \leq 0, d = 1, \dots, D, \tag{3}$$
$$u_{1t}, u_{2t}, v_{1t}, v_{2t} \geq 0,$$

where
D = number of DMUs under evaluation, *t, and d = 1, 2, …, D.*
u_{rt}= coefficient or weight assigned by DMU_t to output *r, r = 1, 2, …, R.*
v_{it} = coefficient or weight assigned by DMU_t to input *i, i = 1, 2, …, S.*

R and *S* denotes the number of inputs and outputs, respectively. In the above model, *R* and *S* are equal to 2. DMU_t is said to be efficient only if $E_t^* = 1$. The model given by (1)-(3) is called an input-oriented model, and E_t^* is called CRS (Constant Scales to Scale) efficiency score (ES).

The cross-efficiency (CE) method consists of self-evaluation (Phase 1) and peer evaluation (Phase 2). The first self-evaluation phase computes ESs using the model by (1)-(3). In the second peer-evaluation phase, the multipliers/weights generated in the self-evaluation phase are applied to all DMUs to get each DMU's cross-efficiency score (CES). Now, the CE for DMU_d is given by

$$E_{td} = \frac{u_{1t}^*EDC_d + u_{2t}^*CDE_d}{v_{1t}^*TLC_d + v_{2t}^*MCD_t}, \qquad t \text{ and } d = 1, \dots, D, \tag{4}$$

DMU_d is a rated DMU, whereas DMU_t is a rating DMU. Zhu (2014) includes self-evaluation value in averaging the appraisals by itself and peers as follows:

$$\bar{E}_t = \frac{1}{D}\sum_{d=1}^{D} E_{dt}. \quad (5)$$

EFFICIENCY-BASED CLUSTERING ANALYSIS (EBCA) METHOD

The proposed EBCA approach, which evaluates each DMU's performance apart from DEA methods, differs from some DEA-based clustering methods that Po et al. (2009) and Chen et al. (2022) propose. As T-DEA classifies all DMUs under evaluation into two groups, separating efficient DMUs from inefficient DMUs, there are two (2) clusters in the proposed method. The first cluster is an efficient one, while the second is an inefficient cluster. Given a set of *D* DMUs with *S* inputs and *R* outputs, all DMUs under evaluation will be classified into one of these two clusters or both for input and output, respectively. What separates the proposed EBCA method from other DEA-based clustering methods is that the global minimum and the maximum points are set to represent the clusters rather than the centroid-based technique. Efficiency or productivity is expressed as the ratio of outputs to inputs. It implies that the greater the efficiency will be as an output increases and/or an input decreases. For the detailed procedure, see Hong (2023).

CASE STUDY WITH OBSERVATIONS

This paper uses major disaster declaration records in South Carolina (SC) for the case study. Forty-six (46) counties are grouped, based on populations and proximity, into twenty (20) counties. By selecting one location from each group, we assume that all population within the grouped county exists in one location. Federal Emergency Management Agency (FEMA) database (FEMA, 2017) reveals that SC has experienced sixteen (16) major natural disaster declarations, such as floods, hurricanes, tornadoes, etc., from 1964 to 2017, and also shows counties where a major disaster was declared. We assume that the county's ERF is disrupted and shut down in case a major disaster is declared. Based on the assumption and historical record, each *AFS*'s risk probability is computed by dividing the years with major weather or natural disasters by the total years. Table 1 lists those risk probabilities. We select the five potential locations for CWHs based on the proportion of area that each site would potentially cover, population, and proximity to Interstate Highways in SC. In most cases, the numbers of CWHs and RDCs to be constructed are pre-specified.

We solve the WGP model in Hong and Jeong (2019) for various values of weight, $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. Each weight changes between 0 and 1 with an increment of 0.1, subject to $\sum_{\kappa=1}^{4} \alpha_\kappa = 1$. The 'Gurobi' Solver Engine of Analytic Solver is applied to solve the model. Using the combinations of the setting of $\boldsymbol{\alpha}$, we have two hundred eighty-six (286) configurations generated by solving the WGP model. We can reduce the 286 configurations to sixty-eight (68) consolidated configurations since several configurations yield the same values of the four performance measures. Each of the 68 configurations representing the optimal locations and allocations of ERFs for a given set of $\boldsymbol{\alpha}$ is considered a DMU.

Table 2 presents the top 25 DMUs ranked by CE-DEA, each performance measure's value, ES for T-DEA, and cross-efficiency scores (CESs), $\boldsymbol{CES_{d|All}}$, when all 68 DMUs are evaluated (**Case 1)**, and the corresponding rankings, $\boldsymbol{R_{d|All}}$. As pointed out earlier, CE-DEA frequently ranks inefficient DMUs higher than efficient DMUs by T-DEA. As shown in Table 2, T-DEA identifies eight inefficient DMUs, $\{DMU_{17}$, DMU_{48}, DMU_{82}, DMU_{84}, DMU_{107}, DMU_{108}, DMU_{157},

DMU_{238}} whose ESs are less than 1.000. But some inefficient eight DMUs are ranked higher than the seventeen efficient DMUs. We observe that *DMU_{125}*, an efficient one by T-DEA, is ranked at the bottom of the top 25 DMUs, while *DMU_{157}*, an inefficient one with an ES of 0.9852, is the highest-ranked one as #8 among the inefficient group. *DMU_{133}*, with the greatest $\boldsymbol{CES_{d|All}}$ of 0.9254, is the top-ranked DMU under **Case 1**.

Table 1: Data for Locations of ERFs

No	City	County	Population (K)	Risk Probability
1	Anderson	Anderson/Oconee/Pickens	373	0.125
2	Beaufort	Beaufort/Jasper	187	0.063
3	Bennettsville	Marlboro/Darlington/Chesterfield	96	0.375
4	Conway	Horry	269	0.375
5	Georgetown	Georgetown/Williamsburg	93	0.438
6	Greenwood	Greenwood/Abbeville	92	0.125
7	Hampton	Hampton/Allendale	33	0.188
8	Lexington	Lexington/Newberry/Saluda	318	0.313
9	McCormick	McCormick/Edgefield	35	0.250
10	Moncks Corner	Berkeley	178	0.313
11	Orangeburg	Orangeburg/Bamberg/Calhoun	123	0.375
12	Rock Hill	York/Chester/Lancaster	321	0.313
13	Spartanburg	Spartanburg/Cherokee/Union	367	0.313
14	Sumter	Sumter/Clarendon/Lee	157	0.375
15	Walterboro	Colleton/Dorchester	135	0.250
16	Aiken†	Aiken/Barnwell	184	0.313
17	Charleston†	Charleston	350	0.250
18	Columbia†	Richland/Fairfield/Kershaw	461	0.375
19	Florence†	Florence/Dillon/Marion	203	0.438
20	Greenville†	Greenville/Laurens	521	0.125

†potential locations for *CWH*

We apply CE-DEA to evaluate these top-25 DMUs only (**Case 2**) and present the CES, $\boldsymbol{CES_{d|\{25\}}}$, and corresponding rankings, $\boldsymbol{R_{d|\{25\}}}$, in Table 2. Consequently, the resulting CES, $\boldsymbol{CES_{d|\{25\}}}$, for **Case 2** is greater than $\boldsymbol{CES_{d|All}}$ for **Case 1** because only the top 25 DMUs are evaluated in calculating $\boldsymbol{CES_{d|\{25\}}}$. The most notable observation is the top-ranked DMU is changed from *DMU_{133}* for **Case 1** to *DMU_{97}* for **Case 2.** In fact, except for two DMUs, *DMU_{108}* and *DMU_{125}*, all DMUs' rankings are changed, demonstrating that the ranks generated by CE-DEA are inconsistent. To denote the change in rankings, the absolute ranking difference (ARD) (=$|\boldsymbol{R_{d|All}} - \boldsymbol{R_{d|\{25\}}}|$), is listed in the last column in Table 2. Surprisingly, *DMU_{28}* becomes the largest one with an ARD of 10, and there are six DMUs with an ARD of 5. We apply the proposed EBCA method to evaluate all 68 DMUs and select the top 25 DMUs. In Tables 3 and 4, the top five DMUs evaluated by CE-DEA and EBCA are presented. Both CE-DEA and EBCA select the same 20 DMUs out of the top 25 DMUs. Now, we depict the two highly-ranked DMUs by both methods, {*DMU_{97}*, *DMU_{133}*}, in Figure 2. Figure 2 shows two sites, {Columbia, Greenville}, are selected as the *CWH* locations by both DMUs. In fact, {Greenville} is selected as the *CWH* location by both DMUs.

Table 2**:** Top 25 Efficient DMUs, Cross Efficiency Score, and Rank Based on DEA Methods

No	DMU #	DEA Method					
		T-DEA	CE-DEA				ARD
		ES	$\boldsymbol{CES_{d\|All}}$	$\boldsymbol{R_{d\|All}}$	$\boldsymbol{CES_{d\|\{25\}}}$	$\boldsymbol{R_{d\|\{25\}}}$	
1	17	0.9469	0.8615	23	0.8987	24	1
2	25	1.0000	0.8867	14	0.9174	19	5
3	26	1.0000	0.8868	13	0.9175	18	5
4	28	1.0000	0.8774	22	0.9290	12	10
5	34	1.0000	0.8806	20	0.9118	21	1
6	35	1.0000	0.8813	18	0.9173	20	2
7	42	1.0000	0.8855	15	0.9328	10	5
8	43	1.0000	0.9010	7	0.9487	6	1
9	48	0.9959	0.8842	16	0.9245	13	3
10	81	1.0000	0.9241	2	0.9535	3	1
11	82	0.9661	0.8872	12	0.9212	14	2
12	84	0.9916	0.8594	24	0.9074	23	1
13	88	1.0000	0.8919	10	0.9211	15	5
14	89	1.0000	0.9229	3	0.9534	4	1
15	91	1.0000	0.8929	9	0.9419	8	1
16	97	1.0000	0.9085	4	0.9547	1	3
17	98	1.0000	0.9069	5	0.9537	2	3
18	107	0.9853	0.8803	21	0.9184	16	5
19	108	0.9969	0.8901	11	0.9296	11	0
20	125	1.0000	0.8468	25	0.8631	25	0
21	133	1.0000	0.9254	1	0.9531	5	4
22	143	1.0000	0.9053	6	0.9469	7	1
23	157	0.9852	0.8939	8	0.9336	9	1
24	180	1.0000	0.8810	19	0.9182	17	2
25	238	0.9618	0.8820	17	0.9095	22	5

$\boldsymbol{CES_{d|All}}$: CES of DMU_d with all DMUs, $\boldsymbol{R_{d|All}}$:Rank for DMU_d for all, $\boldsymbol{CES_{d|\{25\}}}$: CES of DMU_d with top 25 DMUs, $\boldsymbol{R_{d|\{25\}}}$: Rank for DMU_d with top 25 DMUs, **ARD (**Absolute Rank Difference)= $|\boldsymbol{R_{d|All}} - \boldsymbol{R_{d|\{25\}}}|$

Table 3: Top Five DMUs for Each Case by CE-DEA

Rank	**Case 1**: All 68 DMUs under evaluation	**Case 2**: Top-25 DMUs under evaluation
1	DMU_{133}	DMU_{97}
2	DMU_{81}	DMU_{98}
3	DMU_{89}	DMU_{81}
4	DMU_{97}	DMU_{89}
5	DMU_{98}	DMU_{133}

Table 4: Top Five DMUs for Each Case by EBCA

Rank	**Case 1**: All 68 DMUs under evaluation	**Case 2**: Top-25 DMUs under evaluation
1	DMU_{97}	DMU_{97}
2	DMU_{98}	DMU_{98}
3	DMU_{108}	DMU_{108}
4	DMU_{43}	DMU_{107}
5	DMU_{48}	DMU_{48}

COMPUTATIONAL EXPERIENCE

To compare the computational times between CE-DEA and EBCA, we implement the EBCA method in an Excel spreadsheet with VBA (Visual Basic for Applications) on Intel® Xeon ® Gold 5122 HP Z4 Workstation PC (2 processors) with 32.0GB of RAM installed using a 64-bit version of Windows 10 pro for workstations. We randomly generated the values of two inputs and two outputs using a uniform distribution with the minimum and maximum values of all 68 DMUs for the number of DMUs, {68, 100, 200, 300, 400, 500}. A DEA software, *DEAFrontier*, is run for the generated DMUs to find CEs on the same computer. The results are listed in Table 5. Figure 3 depicts the results in Table 5. As expected, Table 8 and Figure 3 show that the computational times for EBCA are almost negligible compared to CE-DEA. It takes less than three seconds for EBCA to get the results with five hundred DMUs, while the DEA software takes almost one thousand seconds to get the CESs. Figure 4 clearly shows that the computational time for CE-DEA sharply increases when the number of DMUs, *D*, increases.

Table 5: Comparison of Computational Time

(Time Unit: Second)

D	68	100	200	300	400	500
EBCA	0.45	0.66	1.12	1.61	2.12	2.67
CE-DEA	45.43	78.48	208.55	421.65	715.63	990.40
Ratio	103.1	118.9	186.2	261.9	337.5	370.9

D: Number of DMUs

Figure 3: Computational time comparison

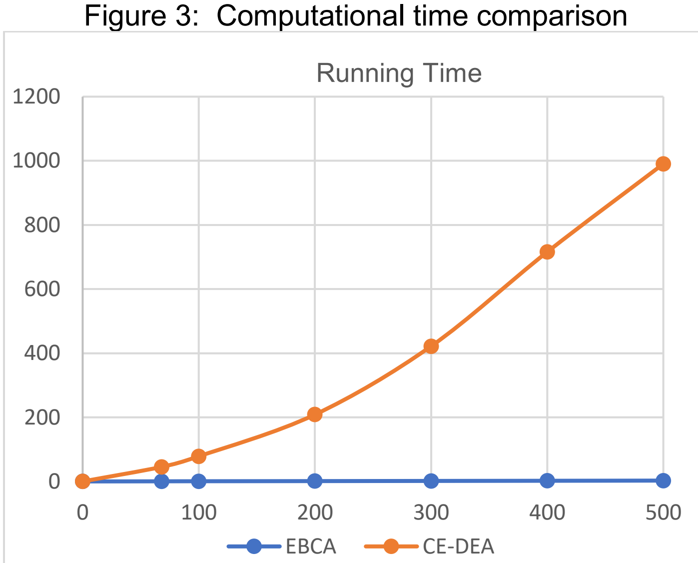
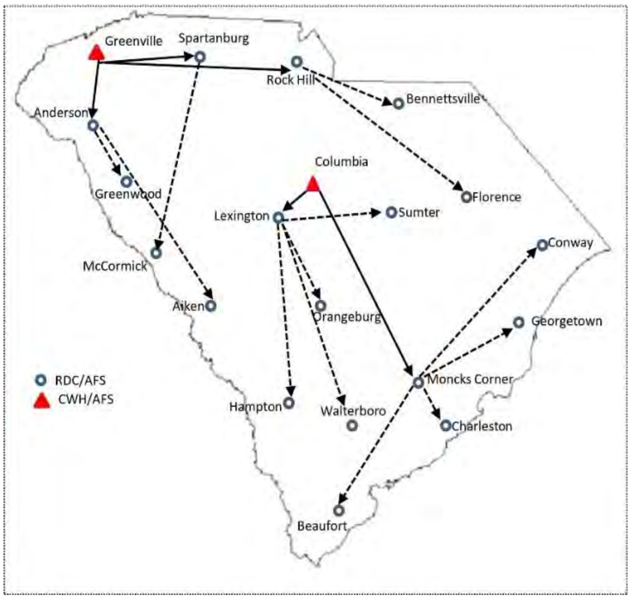
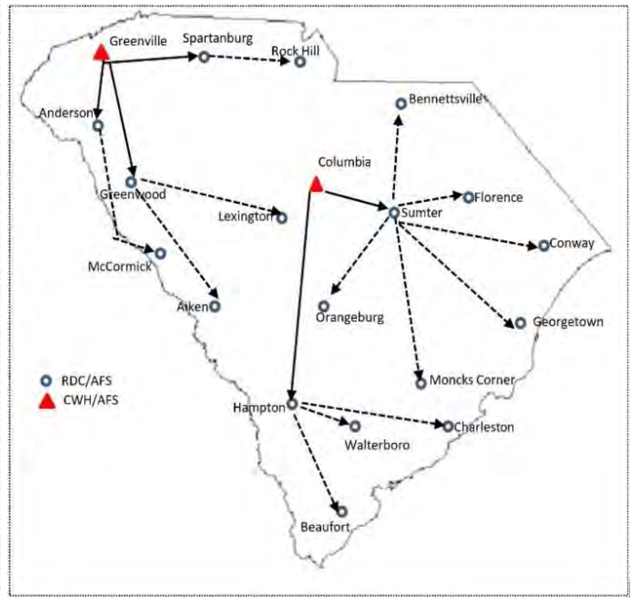


Figure 2: The two efficient emergency logistics network system schemes

DMU$_{133}$ DMU$_{97}$





SUMMARY AND CONCLUSIONS

The design of the emergency logistics network system (*ELNS*) has become an important strategic decision due to the significant damage inflicted by recent natural or human-made disasters, including the pandemic. This paper deals with designing resilient, robust, and efficient *ELNS* so ERFs can distribute relief items to the affected sites at the right time with the right amount of disaster relief items. For planning more balanced *ELNS* schemes, weighted goal programming (WGP) is applied to generate various network alternatives. Some authors used the traditional DEA (T-DEA) method to evaluate these network alternatives. The T-DEA estimates each scheme/DMU on the principle of self-evaluation, ignoring unfavorable inputs/outputs when the efficiency scores are computed. Consequently, the issues related to weak discriminatory power have emerged since multiple DMUs frequently become efficient, and these efficient DMUs are considered equal regarding efficiency scores (ES). The cross-efficiency DEA (CE-DEA) on the principle of peer evaluation was introduced to solve the lack of discriminatory power in T-DEA. But CE-DEA can't eliminate the intrinsic weakness of DEA.

For assessing DMUs more consistently without prejudice, this paper proposes an efficiency-based clustering analysis (EBCA) method, using the overall minimum and maximum values to represent two clusters rather than the centroid of each cluster. Moreover, the proposed method could evaluate and rank the DMUs more flexibly by allowing the decision-maker to assign unequal weights to each input and output.

Using the actual data available for South Carolina, this paper applies the proposed methods and CE-DEA to assess and rank various *ELNS* configurations generated by the WGP model. The proposed EBCA method exhibits its excellent performance over CE-DEA in terms of generating consistent and robust rankings for the DMUs under evaluation. The EBCA approach can be valuable for designing ELNS and other supply chain network systems with multiple outputs and inputs.

The limitation of this study comes from the assumption that if an ERF, either CWH or RDC, is disrupted and can't function, the allocated sites, either RDCs or AFSs, will not be covered. It implies that if a CWH is disrupted, all RDCs assigned to this disrupted CWH and subsequent AFSs won't be covered. To enhance this study, the concept of emergency operation should be considered in case of an ERF shutdown. This paper assumes that only facilities are subject to disruptions, but, in reality, disruption can block the flow of relief items due to route disruptions. Thus, it would be interesting if an emergency backup routing plan is considered for future research.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

An Efficiency-Driven Data Clustering Method for Evaluating Decision-Making Units in the Big Data Context

Jae-Dong Hong
Industrial Engineering
South Carolina State University
Orangeburg, SC 29117
Email: jhong@scsu.edu

ABSTRACT

The data envelopment analysis (DEA)-based methods are viewed as a tool for big data-enabled analytics in efficiency evaluation for the decision-making units (DMUs) with multiple inputs and outputs. However, the critical issue is that excessive running time is required for solving a large set of DMUs in the Big Data context. Another intrinsic issue is that DEA-based models unfairly ignore each DMU's unfavorable inputs and outputs. Consequently, these methods show poor discriminating power, including inconsistent rankings. This paper proposes an efficiency-driven data clustering (EDDC) method to eliminate such critical issues. The proposed method demonstrates its outstanding performance through numerical examples.

KEYWORDS: Data envelopment analysis, Big data-enabled analytics, Decision-making unit, Efficiency-driven clustering method

INTRODUCTION

Among many performance evaluation and benchmarking methods, Data envelopment analysis (DEA) has emerged as the most popular technique that uses Linear Programming (LP) to rate the relative performance of a set of peer decision-making units (DMUs) with multiple inputs and outputs by comparing how well the DMU uses its inputs to produce outputs. The classical DEA (C-DEA), introduced by Charnes et al. (1978), eventually determines which DMUs attain the efficient outcome using given inputs and which do not.

Charles et al. (2021) provide an overview of the current avenues of research for the studies aimed at integrating DEA with Big Data. Zhu (2022) also suggests that DEA should be viewed as a tool or method for data-oriented analytics in performance assessment and benchmarking. However, the DEA-based methods in the Big Data context require excessive running time to evaluate a large set of DMUs, as Barr and Durchholz (1997) show in their paper. In addition, the DEA methods have shown some drawbacks, as mentioned before. The C-DEA frequently produces too many efficient DMUs out of all DMUs under evaluation. In addition, there is no way to rank efficient DMUs since this method can't distinguish them. To remedy this deficiency of C-DEA, Sexton et al. (1986) suggest the cross-evaluation DEA (CE-DEA) method to evaluate and rank DMUs, with the main idea of using C-DEA to do the peer evaluation rather than C-DEA's pure self-evaluation. The CE-DEA method can usually provide a complete ranking of the assessment of DMUs (see Anderson et al., 1993). Due to its enhanced discriminatory power, the CE evaluation has generated a significant number of applications in the DEA literature (see Liang et al., 2008; Wang and Chin, 2010; Gavgani and Zohrehbandian, 2014; Hou et al., 2018; Lee, 2019; Liu et al., 2019;). As Doyle and Green (1994) indicated, the non-uniqueness of CE scores (CESs) often results from alternative optimal weights in the C-DEA model, implying the

CESs depend on the optimization software used. Thus, the non-uniqueness has been criticized as a major drawback of using the CE method as a ranking tool.

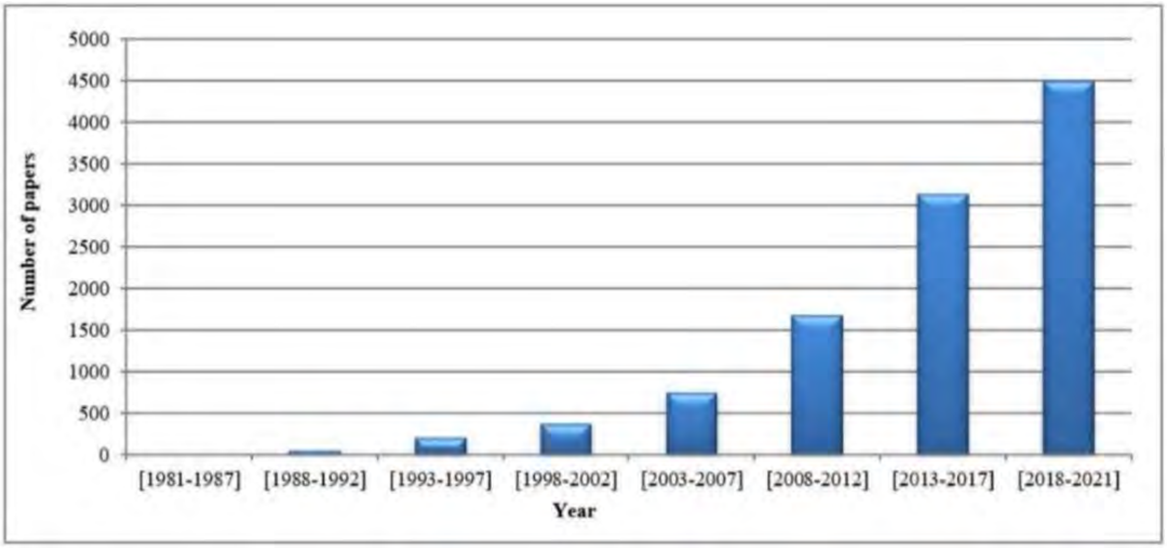
The idea of super-efficiency (SE), mainly developed by Anderson and Peterson (1993), is that a DMU under evaluation is excluded from the reference set of the C-DEA model. Notably, the SE-DEA model has significance for discriminating among efficient DMUs, as Anderson and Peterson (2003) demonstrate. Charnes et al. (1992) use the SE-DEA model to study the sensitivity of the efficiency classification. Deng et al. (2018) and Nayebi and Lotfi (2016) discuss further applications of SE-DEA. The severe issue of using this model is that the adjacent DMUs decide on an efficient DMU's SE score (SES), so it would sometimes be unreasonable or impractical for DMUs to be ranked by the SESs. Thus, each DEA-based method has its own critical issues, drawbacks, or disadvantages, which could become severe as the number of inputs or outputs increases. Frequently, DEA-based models produce inconsistent ESs and unstable rankings for the under-assessment DMUs.

The ranking for a DMU would imply significant meanings since decision-makers can be aware of the current level of a DMU's performance so that they can establish incentive schemes, support policies, and develop strategies to sustain the future business or core competence. No literature has explicitly and seriously discussed this critical weakness of the DEA-based models regarding unstable and unreasonable rankings generated by these models. These weaknesses would be unavoidable because these DEA-based models have unfairly treated unfavorable inputs or outputs. Nevertheless, there has been constant progress in the publications of the DEA from 1978 to 1995. Still, from 1995 onwards, there has been an exponential rise in theoretical development and diverse applications. This trend is reflected in Figure 1 (see Panwar et al., 2022). However, few papers have disclosed the critical weaknesses of DEA-based methods.

Data clustering or cluster analysis is the process of dividing a collection of data instances into groups/clusters. Clustering, which categorizes the data into clusters, aims to identify meaningful segmentations or groupings of objects within a data set. In other words, clustering sorts objects into clusters or groups so that the degree of association is strong among elements/members of the same cluster and weak among elements/members of different clusters (see Delen, 2021). This paper proposes an **efficiency-driven data clustering (EDDC)** method. One of many tools for calculating distance measures is the centroid-based technique. What distinguishes the proposed method from other clustering methods is that the EDDC utilizes the overall minimum and maximum points to represent the clusters rather than the centroid of each cluster. Without ignoring or sacrificing any given inputs/outputs, the EDDC method would eliminate the critical issues of DEA-based evaluation models and rate the DMUs more impartially and consistently than the DEA-based methods.

As Big Data research becomes an important area of operations analytics, DEA is evolving into data-enabled analytics and a data-oriented data science tool for various operational analytics. Various researchers show that Big Data can be applied to improve company productivity or efficiency and will be essential for enterprises to grow and achieve a competitive advantage. When researchers attempt to integrate DEA into Big Data, the serious barrier between Big Data and DEA is the excessive running time for solving a large set of DMUs in the Big Data context. The proposed EDDC process takes advantage of the simple computation, not requiring any optimization software that all DEA-based models need. So, the proposed EDDC method would assist practitioners and researchers in performing a more refined and consistent efficiency evaluation than the DEA-based methods. Besides, the proposed procedure would quickly provide a benchmarking framework for DMUs to improve efficiency.

Figure 1: Year-wise publication of DEA papers (Panwar et al., 2022)



Some researchers apply a DEA-based approach for clustering. In other words, they build clusters for each DMU based on the results of DEA rather than a distance measure. For example, Po et al. (2009) propose a DEA-based clustering approach rather than a distance measure. Their proposed method employs the piecewise production function derived from the C-DEA method to group the input and output items. Following Po et al. (2009), Chen et al. (2022) propose a clustering approach within the CE-DEA framework to obtain a clustering result based on the production relationship between the inputs and outputs date of DMUs. These DEA-based clustering methods still carry DEA's intrinsic weakness of ignoring unfavorable inputs and outputs.

DATA ENVELOPMENT ANALYSIS-BASED MODELS

The ratio of weighted outputs to weighted inputs for measuring the relative efficiency of DMU_j is expressed as an objective function for the fractional DEA model. The fractional DEA model (Cooper et al., 2011) is stated as follows:

Objective Function: Maximize the efficiency rating θ for DMU_j

$$Max\ \theta_j = \frac{\sum_{r=1}^{s} u_{rj} y_{rj}}{\sum_{i=1}^{m} v_{ij} x_{ij}}, \quad j = 1, 2, \dots n. \tag{1}$$

The objective function in (1) is subject to the constraint that when the same set of u and v weights (multipliers) is applied to all other DMUs to be compared, no DMU will be more than 100% efficient as follows:

$$\frac{\sum_{r=1}^{s} u_{rj} y_{rw}}{\sum_{i=1}^{m} v_{ij} x_{iw}} \leq 1, \forall j \ \&\ w = 1, 2, \dots, n. \tag{2}$$

$$u_{rj}, v_{ij} \geq 0, r = 1, \dots, s; i = 1, \dots, m; w = 1, 2, \dots, n,$$

where

j = DMU j being evaluated in the DEA analysis, j = 1,…, n
y_{rj} = amount of output r produced by DMU_j
x_{ij} = amount of input i consumed by DMU_j
n = number of DMUs under evaluation
m = number of inputs used by DMUs
s = number of outputs generated by DMUs
u_{rj} = multipliers or weight assigned by DEA to output r for DMU_j
v_{ij} = multipliers or weight assigned by DEA to input i for DMU_j

Based on the fractional DEA model in (1)-(2), the traditional (T), cross-efficiency (CE), and super-efficiency (SE) DEA models were developed. The C-DEA model is formulated as the following LP problem, where E_{jj} represent the efficiency score (ES) for DMU_j:

$$max \quad E_{jj} = \sum_{r=1}^{s} u_{rj} y_{rj}, \tag{3}$$

subject to

$$\sum_{i=1}^{m} v_{ij} x_{ij} = 1, \tag{4}$$

$$\sum_{r=1}^{s} u_{rw} y_{rw} - \sum_{i=1}^{m} v_{ij} x_{iw} \leq 0, w = 1, \dots, n, \tag{5}$$

$$u_{rj}, v_{ij} \geq 0, r = 1, \dots, s; i = 1, \dots, m.$$

The CE-DEA method, consisting of two phases, was proposed to rank DMUs with the central idea of applying DEA to do peer evaluation rather than pure self-evaluation (see Sexton et al., 1986). The weights or multipliers from the first phase are applied to all DMUs to get each DMU's cross-efficiency score (CES) in the second phase. In the first phase, the above LP model in (3)-(5) is solved to find the ES of DMU_j. To denote the peer evaluation, let E_{jw} represent the DEA score for the rated DMU_w, w = 1, 2, …, n, using the optimal weights /multipliers that a rating DMU_j has chosen in the model (3)-(5). Now, E_{jw} is given by

$$E_{jw} = \frac{\sum_{r=1}^{s} u_{rj}^{*}\, y_{rw}}{\sum_{i=1}^{m} v_{ij}^{*}\, x_{iw}}, \qquad j \ and \ w = 1, \dots, n. \tag{6}$$

Then, the CE score for DMU_w is defined as follows:

$$CE_w = \frac{1}{n} \sum_{j=1}^{n} E_{jw} \tag{7}$$

As mentioned before, the CE-DEA method of ranking the DMUs with many inputs and outputs shows inconsistent results.

The SE-DEA would generate a super-efficiency score (SES) obtained from the conventional DEA model after a DMU under evaluation is excluded from the reference set (Anderson and Peterson, 1993). In the SE method, the frontier line generated from the remaining DMUs changes for each efficient DMU to be evaluated, so the SESs of efficient DMUs can have higher values than 1, which is the maximum value in ES obtained by other DEA methods. The SE-DEA model, which has been applied significantly for ranking efficient DMUs, is given by

$$Max \quad SES_j = \sum_{r=1}^{s} u_{rj} y_{rj}, \tag{8}$$

subject to

$$\sum_{i=1}^{m} v_{ij} x_{ij} = 1, \tag{9}$$

$$\sum_{r=1}^{s} u_{rw} y_{rw} - \sum_{i=1}^{m} v_{ij} x_{iw} \leq 0, w \neq j, w = 1, \ldots, n, \tag{10}$$

$$u_{rj}, v_{ij} \geq 0, r = 1, \ldots, s; i = 1, \ldots, m.$$

EFFICIENCY-DRIVEN DATA CLUSTERING (EDDC) METHOD

The proposed EDDC approach, which evaluates each DMU's performance without applying DEA methods, differs from some DEA-based clustering methods proposed by Po et al. (2009) and Chen et al. (2022). As C-DEA can separate efficient DMUs from inefficient DMUs, only two (2) clusters are in the proposed method. The first cluster is an efficient one, while the second is an inefficient cluster. Given a set of *n* DMUs with *m* inputs and *s* outputs, all DMUs under evaluation will be classified into one of these clusters or both for input and output, respectively. What separates the EDDC method from other DEA-based clustering methods is that the global minimum and the maximum points are set to represent the clusters rather than the centroid-based technique. Efficiency or productivity is expressed as the ratio of outputs to inputs. It implies that the greater the efficiency will be as an output increases and/or an input decreases. See Hong (2023) for detailed procedures.

NUMERICAL EXAMPLE

This paper considers the numerical example Liang et al. (2008) illustrate with five DMUs. The data for each DMU consists of three inputs (x_{1j}, x_{2j}, x_{3j}) and two outputs (y_{1j}, y_{2j}), as shown in Table 1. CESs by CE-DEA, SCES by SE-DEA, and ES_{jN}^{C} by the EDDC method, along with the corresponding ranks, are reported in Table 2. Table 2 exhibits that all methods, including the proposed EDDC, rank the two efficient DMU_3 and DMU_2 as #1 and #2 consistently, but each method ranks the other inefficient DMU_1, DMU_4, and DMU_5 differently.

For further investigation, the three inefficient DMUs, after excluding two efficient DMUs, are evaluated and reported in Table 3. Since the two efficient DMUs are excluded, C-DEA rates these three DMUs as efficient with an ES of 1. Table 3 shows the ranks and the expected ranks, denoted by [R], based on the ranks in Table 2. We observe from Table 3 that SE-DEA and EDDC rank these DMUs consistently, but CE-DEA does not. Table 3 also shows that two DEA-based models, CE- and SE-DEA, rank DMU_4 higher than DMU_1 or DMU_5, whereas the EDDC ranks DMU_5 higher than DMU_4 and DMU_1.

Table 1: Five Decision-Making Units

DMU	Input			Output	
	x_{1j}	x_{2j}	x_{3j}	y_{1j}	y_{2j}
1	7.0	7.0	7.0	4.0	4.0
2	5.0	9.0	7.0	7.0	7.0
3	4.0	6.0	5.0	5.0	7.0
4	5.0	9.0	8.0	6.0	2.0
5	6.0	8.0	5.0	3.0	6.0

Table 2: Comparison of Diverse Efficiency Scores and Rankings for Five DMUs

DMU	C-DEA		CE-DEA		SE-DEA		EDDC	
	ES	R	CES	R	SES	R	ES_{jN}^{C}	R
1	0.6857	5	0.5191	4	0.6857	5	0.3278	4
2*	1.0000*	1	0.9161	2	1.1200	2	0.9716	2
3*	1.0000*	1	0.9571	1	1.5000	1	1.0000	1
4	0.8571	3	0.6985	3	0.8571	3	0.1698	5
5	0.8571	3	0.4942	5	0.8571	3	0.4576	3

*: Efficient by C-DEA; R: Rank

Table 3: Comparison of Efficiency Scores and Rankings for Three Inefficient DMUs

DMU	C-DEA		CE-DEA			SE-DEA			EDDC		
	ES	R	CES	R	[R]	SES	R	[R]	ES_{jN}^{C}	R	[R]
1	1.0000	1	0.7714	3	2	1.1238	3	3	0.9597	2	2
4	1.0000	1	1.0000	1	1	2.1000	1	1	0.3466	3	3
5	1.0000	1	0.8055	2	3	2.1000	1	1	1.0000	1	1

R: Rank; [R]: Expected Rank

As DEA models implicitly assume equal weight to each input and output, it is unreasonable for two DEA-based models to rank DMU_4 higher than DMU_5. DMU_4 and DMU_5 have the same minimum input, 5.0, and the maximum output, 6.0. But DMU_4 has the other two inputs, 9.0 and 8.0, which are greater than or equal to those of DMU_5, 6.0, and 8.0. Except for the same maximum output value of 6.0, DMU_5 has a higher output value of 3.0 than 2.0 for DMU_4, implying that DMU_5 should be ranked higher than DMU_4. Thus, **the ranks generated by the EDDC method would be more rational and logical than those by DEA-based models.**

We implement the EDDC method in an Excel spreadsheet with VBA (Visual Basic for Applications) on Intel® Xeon ® Gold 5122 HP Z4 Workstation PC (2 processors) with 32GB of RAM installed using a 64-bit version of Windows 10. We randomly generated the values of three inputs and two outputs using a uniform distribution with the minimum and maximum values from Table 1 for the numbers of DMUs, {15, 50, 75, 100, 150, 200, 300, 400, 500}. A DEA software, *DEAFrontier*, using an Excel spreadsheet, is run for the generated DMUs to find CES on the same computer to compare the running times between the DEA-based and proposed EDDC methods. The results are listed in Table 4, including the ratio of running time for CE-DEA to that of EDDC. Figure 2 depicts the results in Table 4. As expected, we see from Table 4 that the running times for EDDC are almost negligible compared to CE-DEA. Figure 2 clearly shows that the running time for CE-DEA sharply increases when the number of DMUs increases.

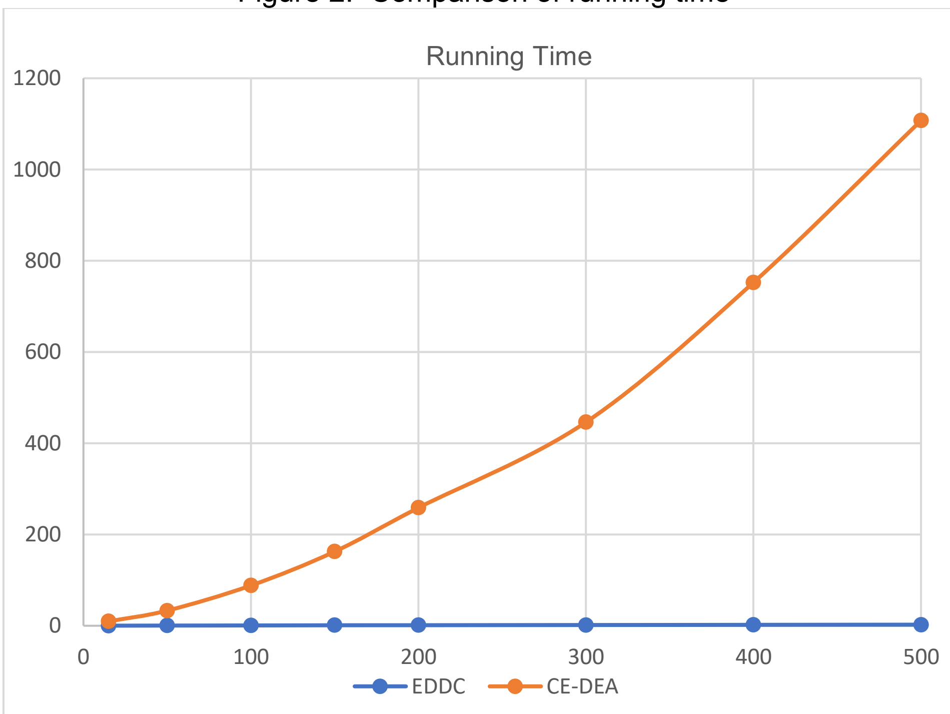
Table 4: Comparison of Running Time.

(Time unit: second)

n	15	50	100	150	200	300	400	500
EDDC	0.14	0.38	0.61	1.02	1.18	1.57	1.95	2.31
CE-DEA	9.5	33.1	88.1	162.5	258.6	446.2	752.4	1,107
Ratio	66.0	86.7	145.7	159.5	219.2	284.9	385.2	478.8

n: Numbers of DMUs

Figure 2: Comparison of running time



SUMMARY AND CONCLUSIONS

Various DEA methods enable decision-makers to measure the efficiency of DMUs and rank them based on efficiency scores (ESs). Ranking DMUs based on the ESs generated by the C-DEA method shows a significant drawback due to the self-evaluation principles. Several ranking methods based on the C-DEA have been proposed, but no ranking method has been found to be either a universal or superior method for ranking DMUs. The researchers, who developed various DEA-based ranking methods, have asserted that the absence of global assessment criteria makes evaluating all the presented methods reviewed by their papers impossible. They conclude that each method could be better than others according to the decision maker's preferences and evaluation objectives, depending on the evaluation's nature (see Aldamak and Zolfaghari, 2017).

The DEA-based approaches' severe weakness is their biased preference for specific inputs and/or outputs. All DMUs under evaluation can only use favorable inputs and/or outputs to boost their own efficiency scores, dropping the unfavorable inputs and/or outputs. For the purpose of evaluating DMUs more consistently without prejudice, this paper proposes an efficiency-driven data clustering (EDDC) method, using the overall minimum and maximum points to represent two clusters rather than the centroid of each cluster. Using the two well-known examples in the DEA textbook (Cooper et al., 2007) with either input or output fixed at a given value, the EDDC method evaluates and ranks the DMUs more reasonably and consistently than DEA-based methods. In addition, the proposed method could assess and rank the DMUs more flexibly by allowing the decision-maker to assign unequal weights to each input and output.

We also apply the proposed method to evaluate a well-known numerical example, which has been considered by several authors, to compare the DEA-based methods. This numerical example shows that the rankings generated by the DEA-based methods show such a significant weakness, especially for the top-notch DMUs. When the top-notch DMUs are evaluated, the DEA-based methods allow a previously lower-ranked DMU to take over the top-ranked DMUs. By contrast, the rankings generated by applying the EDDC do not change when some lower-ranked or inefficient DMUs are removed from evaluation.

The EDDC method can quickly get results, as shown in Table 4. Contrary to the DEA-based methods, the proposed method does not require any optimization software to evaluate DMUs under evaluation. The results and observations through the numerical examples demonstrate that the EDDC method works well and would be considered an appropriate tool for evaluating a large set of DMUs in the Big Data context.

ACKNOWLEDGMENTS

This material is based upon work that is supported by the U.S. Department of Education, Grant No. P382G170106.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

An Emerging Adult Patient Portal Behavioral Model

Navya Velverthi
University of North Texas- Denton
Email: navyavelverthi@gmail.com

Victor Prybutok
University of North Texas
Email: Prybutok@unt.edu

Lingzi Hong
University of North Texas- Denton
Email: lingzi.hong@unt.edu

ABSTRACT

The rise in health information technology has provided patients with patient portals for better healthcare management. These portals enhance patient engagement and satisfaction, yet their adoption is limited, especially among emerging adults beginning self-management. This study, rooted in the Technology Acceptance Model and Perceived Risk Theory, examines emerging adults' behaviors toward these patient portals. Data from 18–29-year-olds was analyzed using structural equation modeling. Findings highlight the need for creating awareness and training programs. Such initiatives can minimize patients' perceived risks and promote portal utility. Results also show personal innovativeness affects intentions to adopt and use patient portals.

KEYWORDS: Patient Portal, Technology acceptance model, Emerging Adults, Perceived Risk Theory, and Adoption and Use.

DECISION SCIENCES INSTITUTE
Analysing containers' truck productivity using data analytics

Ma, Nang Laik
Singapore University of Social Sciences
Email: nlma@suss.edu.sg

ABSTRACT

This paper focuses on the impact of container truck movement and activities at yard operations at one of the busiest container terminals in Asia. The long queuing time has become a severe issue for efficient flow in container terminal operations. After the initial exploration of data and insight sharing, we developed two predictive models to predict whether the truck is productive. From the models, the authors also determine the critical factors that affect the container truck productivity index. Finally, two predictive models were compared: the logistics regression model and the decision tree model.

KEYWORDS: data analysis, predictive model, Container terminal, yard operation, truck, predictive models, productivity

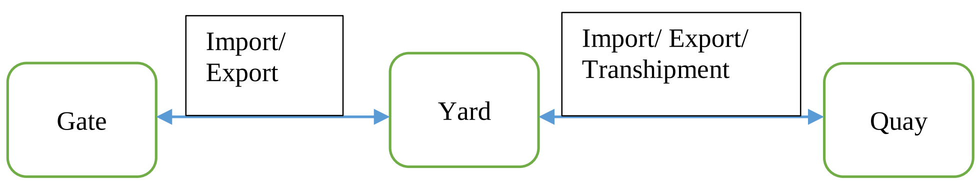
1. INTRODUCTION

Today, 95% of international cargo are transported through container terminal via sea transport, especially with the recent booming of e-commerce. Containers are measured in terms of TEU (twenty-footer equivalent unit). Moving goods across borders via sea transport due to its low cost has become a norm for non-perishable goods. Therefore, the container terminal operations' efficiency is essential to its business success. Containers throughput in some of the busiest terminals in Asia is millions of TEUs annually, thus container terminals are imperative, and their efficiency will affect the whole supply chain. Most container terminals will invest a large amount of money in information technology and advanced equipment to load or unload the containers to shorten the vessel turnaround time at the container terminal. There is also a seamless data transfer between the shipping lines and the container terminal to make the documentation easy and transparent. Operations Research (OR) methods have been used in the planning systems of the container terminal to minimize traveling time, reduce congestion within the port, and improve the productivity of yard and ship operations. This paper's focus will be on the analysis of container truck productivity.

Container terminal operation is essential to the world economy. In the year 2020, the world container throughput reached 797 million TEUs, according to Statista. It is equivalent to an annual increase of 3% from 2012 to 2020. Today, the vessel's size is also enormous. The largest container ship has a capacity of 24,000 TEU. As a container terminal, it faces many challenges in providing exemplary service to the shipping lines to minimize the vessel port stay. Our client, one of the busiest container terminals in Asia, strongly believes in investing in information technology, software development, and optimization engines to automate some of the mandate processes and make better decisions with the aid of artificial intelligence and a decision support system. Unfortunately, due to the NDA signed, we cannot share the real-world operations data, and some of the results shown in this paper have been masked to protect the client's interest. However, this idea shared in this paper will still be helpful to readers who would like to know the ongoing research and development related to container terminals and how data analytics can help to identify the pinpoint in the business process.

A port has three critical areas: gate, yard blocks, and quay where the ships are berthed. Import containers are discharged from the ship, temporarily stored in the yard, and collected by the agent for local consumption. The reverse flow is true for export containers that come to the terminal via the gate to be loaded onto another loading vessel for a foreign destination. Finally, transhipment containers are unloaded from the discharging ship, assigned to a yard location, and loaded onto another vessel to transport them to their destination. The graphic representation of container flows in a typical container terminal is shown in Figure 1 below.

Figure 1: Various container flows in the container terminal



Here is a detailed explanation of the different types of containers within the container terminal.

- **Import:** This is getting containers from ships to other countries through the port for local consumption. For example, a ship may carry the finished products from China and transport them to the USA as import containers.
- **Export:** The containers due for the USA from China are export containers in China's port.
- **Transhipment:** Containers arriving from the source location to the port will be unloaded and stored in the yard, unlike the import containers. When the loading vessel comes, the transhipment containers will be loaded onto the ship for the destination. For example, a transhipment container from Vietnam may arrive in Singapore port and later be transported to the USA by sea.

On the berth side, upon the vessel's arrival, the import or transhipment containers from the ship will be unloaded by the Quay Cranes (QC) onto the container trucks (CT). CT will then transport the container to the destination yard location. Yard Cranes (YCs) are gigantic pieces of yard equipment used to unload or load containers from the container truck and store them on the yard stack. Due to their size, YCs can move up and down the yard block, and they usually can handle 20 to 30 containers per hour. Our paper will only focus on Rubber Tired Gantry Cranes (RTG) and Rail Mounted Gantry Cranes (RMG). Using the planning data available, the yard manager would estimate the number of YCs required with the help of the yard planning system to load/unload containers during operation time. Some trucks can come to the yard via the gate anytime within a pre-defined time window that the drivers booked. These trucks will also require service from YC to unload the containers.

YC will unload the containers from CT and stack them in the yard. Terminals will store the transhipment containers from the yard. The import containers will later be moved out of the container terminal by the shipping line or agents for local consumption within the country. The movement of containers from berth to yard, yard to berth, or other terminals is carried out by CT. The role of the container terminal operator is to ensure that the cargo can be loaded on time without any damage or delay. With so many containers being moved around the terminals daily, terminal managers must ensure that CT is efficiently scheduled with the minimum waiting time.

In this paper, we analyzed the operational data from one of the container terminals in Asia, focusing on the yard activity and productivity of the CT. In the following few sections, we will explore the one week of historical data, determine the productivity index for CT throughout the day, and select the significant factors impacting CT productivity. Furthermore, we propose to develop a predictive model to predict if the CT is productive. Finally, we will measure the accuracy rate of our predictive models, followed by the conclusion and future research direction.

2. LITERATURE REVIEW

Container terminals are a vital part of the supply chain to transport bulky goods worldwide at the lowest cost. Thus, a lot of research is done in the areas of ship operation, yard operations, risk management, and vehicle routing problem within the container terminals. There are three essential review papers have been published. Vis and de Koster (2003) describe a wide range of business problems faced by the container terminal and reviews the various quantitative models used to solve them. It provides an excellent general overview of container terminal operations problems. In Stahlbock and Voß (2008), the authors gave an updated literature review of the container terminal problem as the first review paper. Finally, another paper by Steenken et al. (2004) provides an overview of operations, supply chain, and logistics related to the container terminal.

Most container terminals will plan the ship as an independent entity. However, this approach may conflict if the boats are berthing nearby at the quay as they need to compete for resources such as QC and CT. Gambardella et al. (2001) proposed a solution to solve the resource allocation and scheduling problems related to loading and unloading operations in the terminal. At the macro level, the model determines the number of QC required to load or unload containers at the quayside and the number of YC needed to store those containers in the yard. At the next level, a scheduling problem is formulated to compute containers' loading and unloading time for each of the allocated cranes. The problem is solved efficiently using branch and bound search, and results are validated using the simulation model.

Before the arrival of the vessel, as part of the planning process, the berth planers need to allocate a berth for all the incoming ships for a day using Gantt Charts, which are suitable for visualization. For the berth planners, they need to ensure that there is a safe distance between different vessels when they allocate berths along the quayside. There should also be no overlap for the berth allocation as the vessel may delay due to unforeseen circumstances. When the yard planners plan for the container locations in the yard, they also need to consider the berth location to ensure that the traveling time between the berth and the yard is minimized. A few researchers focus on berth planning. Guan and Cheung (2004) consider the berth allocation problem for the vessel to minimize the total weighted flow time of the vessel. Two mathematical formulations have been developed to shorten the computation running time, and the results show that running the model in real time is efficient. The first model is to create a tree search heuristic, while the other is to develop a lower bound to speed up the search procedure. The computation result shows that the proposed method is superior to standard packages like CPLEX. Every vessel will require a berth location in the wharf or quay during its port stay. In the paper, Kim and Moon (2003) formulated a mixed integer linear programming model for the berth scheduling problem. They used simulated annealing to solve the berthing scheduling problem, and the results showed that the heuristics method produces near-optimal solutions.

Load sequencing determines the order in which the outbound containers, such as export containers, should be loaded onto the vessel from the yard. Kim et al. (2004) developed a beam algorithm to solve the load sequencing problem in the port with the consideration of their yard location. Due to the complexity of the problem, it has been broken down into two subproblems. The pickup schedule was first constructed with the number of containers to be transported at each yard block, and in the second problem, the research determined the load sequence. Numerical results show that a beam search algorithm performs better than an ant colony algorithm.

Chen et al. (2007) focus on the integrated scheduling problems of the quay crane, yard crane, and trucks to transport the containers between the yard and the quay. The authors consider all the unloading and loading tasks sequentially and propose using the tabu search approach. With a good initial solution, tabus search results yield good performance results.

Hong Kong handles million TEU of containers annually. Chung et al. (2002) used real-world data from one of the container terminals in HK to solve the yard cranes scheduling problem. The authors wanted to maintain an efficient container workload across different yard blocks. The Mixed Inter Programming (MIP) model has been developed to solve this problem, and the model's objective is to minimize the unfinished workload at the end of the planning horizon. Due to the long computation time to get the optimal answer from the exact solution, Lagrangian decomposition methods have been proposed. The proposed model can solve the real-world practical problem within the specific time frame and thus can be used for the dynamic situation in the terminal.

With the increasing global trade, there is a need for efficient operations in the container terminal. Das and Spasovic (2003) focus on scheduling straddle carriers, which are used to move the containers in the sea port to minimize empty travel and delay customer service. The methods proposed will dynamically assign the jobs to the straddle carriers, which become available. The authors used the developed method and tested it in the largest container terminal in Port of New York and New Jersey. A simulation model showed that the proposed model performs better than two other alternative strategies.

Hadjiconstantinou & Ma (2008) first developed a decision support system to determine the optimal number of straddle carriers required in the container terminal by minimizing the overall storage and handling cost. The authors also used a discrete event simulation model for the container terminal operations. They tested the solution given by the optimization model and robustness of the straddle carrier deployment policy in one of the largest container terminals in Greece – the Port of Piraeus. Similarly, Ma & Hadjiconstantinou (2008) proposed a unified approach to simultaneously combine container assignment and yard crane deployment problems within a specific time horizon. The proposed model helps determine the optimal flow of containers in the yard.

Finally, Zhang et al. (2003) are interested in efficient container storage at the container terminal in Hong Kong. They have included all the resources available in the sea port, such as QC, YC, yard storage space, and a truck, which will transport the containers from the quayside to the yard side and vice versa. To solve the problem, the authors considered a rolling time horizon; the first problem is allocating the containers in the yard of balancing the workload among different yard blocks. In the second level, the containers in each vessel are considered to minimize the total traveling distance to transport the containers from the yard location to the berth. The authors obtained the computational results within a short timeframe, and the approach reduces the workload imbalance in the yard and makes the terminal operations very efficient.

From the literature review, we identified gaps in the research. More litter research needs to be done to analyze the productivity index of container trucks in the container terminal. Since container terminal operations or planning problems are very complex, usually NP-hard problems, heuristics methods such as tabu search, beam search, or decomposition methods have been used to solve the real-world problem within a reasonable timeframe. In addition, most of the research is based on optimization methods, and few predictive models have been used in the area of the container terminal. Our contribution is two folds. Firstly, we focus on the analysis of operational data to understand the productivity index of CT at the container terminal throughout 24 hours. Secondly, using the insights from the data analysis, we would like to harness the power of data analytics to develop a predictive model to forecast if CT is productive. We will share our method and approach in the following few sections.

3. BUSINESS PROBLEM AND DATA EXPLORATION

To analyze the CT productivity, we first define the productivity of CT as the number of containers being handled divided by the total time taken to service all these containers by the CT. It includes traveling time plus waiting time plus unproductive move time. Since there is less variation in the no. of containers being handled by CT, thus our focus will be on reducing the denominators. To reduce the total time taken, we should reduce the waiting time, traveling time, and unproductive time.

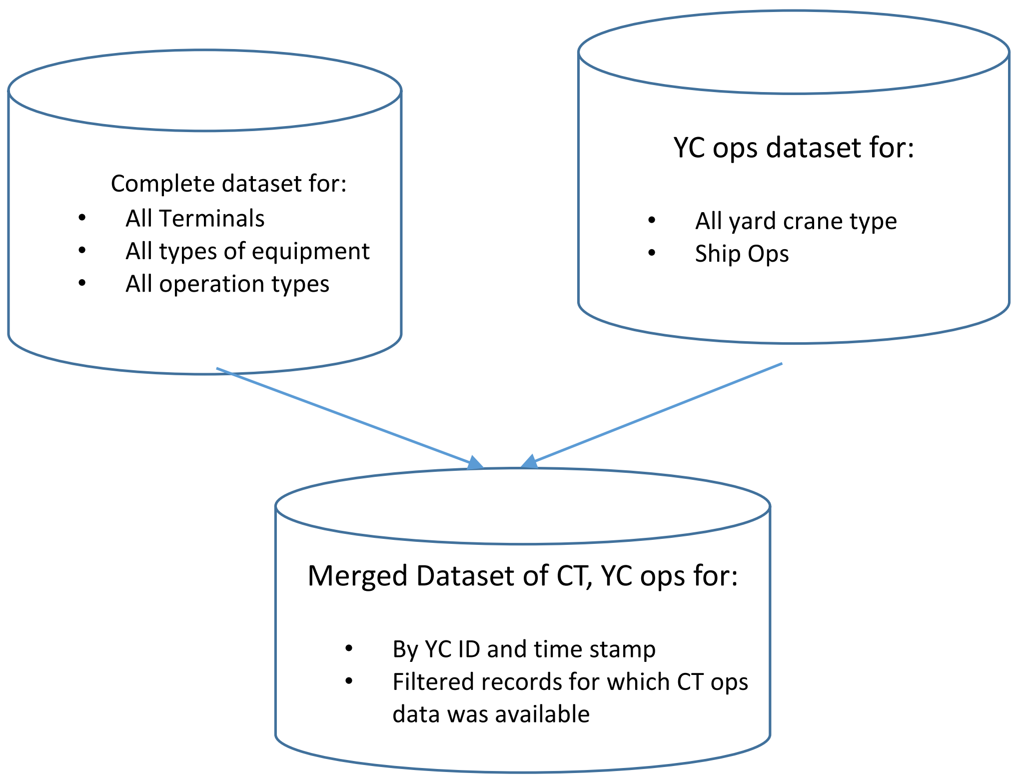
$$CT\ productivity = \frac{Total\ number\ of\ containers\ handled}{Total\ time\ (travel\ time + waiting\ time + unproductive\ time)}$$

CT productivity is also measured with respect to the operations it performs. These are some of the functions in which CT will be involved.

- Ship Ops: Movement of the container from quay cranes to yard or from yard cranes to quay cranes.
- Shifting: Movement of the container from one block of the yard to another block. Usually shifting jobs are unproductive move and should be avoided.
- Gate Ops: Movement of export containers from yard to gate and import containers from the yard to the gate.

Datasets from various operating systems, such as ship operations, yard operations, and CT operations, will be used. CT operation data include records of each trip made by CT and the activities involved. The data are loaded to MySQL database and created a joined table with both the datasets from yard and ship operations. As the opeartional datasets are structured except for a few columns, which required transformation. We have used one-week operational data from the yard and ship operations database, and there are a few hundred thousand records for our analysis.

Figure 2: Overview of data flow



The initial analysis aimed to understand how the productivity of CT varies based on above mentioned parameters like terminal, yard area, equipment type, and ship operations. The research showed below results:

- The yard was the main area with very low productivity compared to the berth. It is due to a lot of activities happening in the container yard areas compared to the quay or wharf side.
- Regarding equipment, Yard Cranes lead to low productivity compared to Quay Cranes. The ratio between QC and YC is 1:2, meaning we need two YCs to support one QC. For example, if the QC rate is 40 moves per hour, the YC rate is about 20 moves per hour.
- CT productivity also varies for a different types of container movements.

Based on the initial finding, we will create a flag or binary variables such as isLoading, and isShifting to identify if the task involves loading or shifting jobs. We need to perform a lot of data transformation for all these categorical variables to binary input values during the data preparation stage before we develop the predictive model.

4. MODEL DEVELOPMENT & EVALUATION

Firstly, we created a dynamic bar graph to show the productivity index for Ship Ops. Then, the bar chart is active and dynamic, where we can view the productivity for the below parameters, which are shown in a list box.

- Terminals (T1, T2, T3)
- Activity Hours (Hour 00 to 23)
- Various Yard Equipment Types (e.g., RMG, RTG)
- Contractor Details

For each parameter, CT Productivity is shown. The user can select any specific parameter to see the results specific to that parameter. The actual values of each productivity bar are shown separately in a pivot table where the user can compare different input values and output.

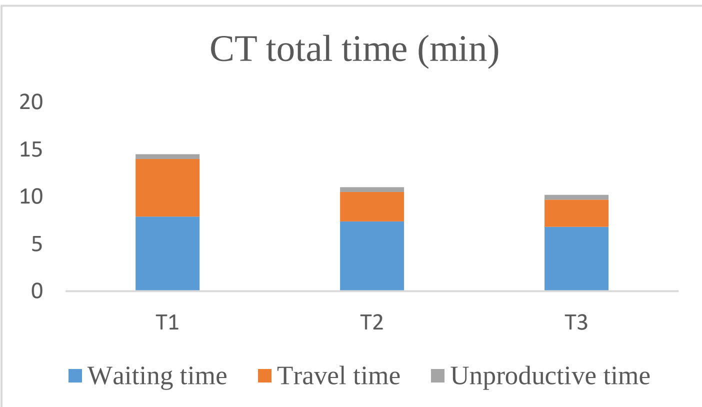
Figure 3: CT productivity across 24 hours

From the figure above, productivity is lower at the 7th hour, followed by the 19th. Finally, productivity is highest at midnight and the 21st hour. CT productivity is defined as the total number of boxes handled divided by the trip time taken. Since there cannot be many variations in the no. of containers processed by a CT as we are focusing only on the productive moves, total time is the main factor due to which productivity varies.

To better understand the total time variation, another bar graph is plotted to show the average travel time, waiting time, and average unproductive time of CT in the terminal.

Table 1: CT activity time			
Terminal	Waiting time (min)	Travel time (min)	Unproductive time (min)
T1	7.9	6.1	0.5
T2	7.4	3.1	0.5
T3	6.8	2.9	0.5

Figure 4: CT travel time, waiting time, and unproductive time



The above Table 1 and Figure **4** illustrate that the traveling time is double in T1 as compared to the other two terminals. More efficient container storage space planning is required in T1. The waiting time of CT in T1 is also the highest in T1 as compared to the other two terminals. T3 has the lowest traveling and waiting time among the three terminals. If the waiting and traveling time is high, then productivity is low.

Insights: For each parameter, the CT Productivity varies, and below are some observations:

- Terminals: Terminal T1 has the lowest productivity among all the port.
- Equipment: Out of RTG and RMG, RMG shows the lowest productivity.

By mapping these observations with the travel time details graph, we can conclude that the T1 terminal has the lowest productivity due to long travel and waiting time. Similarly, Since T3 has the most minor waiting and traveling time, the CT productivity at T3 is much better than the other two terminals. Furthermore, we also analyzed the waiting time for CT, which is more than 15 minutes (>15 min) and less than or equal to 15 minutes (<=15 min). From the analysis, about 10% of the CT will have to wait more than 15 minutes in the yard. The congestion hours are during the 14^{th} to 15^{th} and 22^{nd} to 23^{rd} hours of the day. Congestion typically occurs when CT is queuing for YC to be served and YC is busy helping other CT. Queuing can be further drilled down to the lack of a yard crane in the yard block, and YC was busy loading or unloading other trucks.

The above analysis was a passive analysis that might help understand the high-level issues in CT productivity and provide details on when and where the productivity is low. But the visual analysis needs to determine the significant factors with the priority of its significance for low productivity. We embark on the data analytics journey to develop a predictive model to forecast CT productivity.

Predictive modeling, such as regression or decision trees, are models created to forecast the outcome or predict the result of the target variable, which can be a numeric value or categorical variable. The model is made up of several predictors, which are factors that are likely to influence future behavior or results. In this paper, we want to predict CT's productivity using two regression models: Decision Tree (DT) and Logistics Regression(LR). The former will identify significant factors which highly influence CT productivity. A decision tree is one of the most popular data mining models as it is explainable and easy to understand the rule as What-if-else. The latter will prioritize the level of significance for these factors. The prediction modeling is done in SAS Enterprise Miner.

To do the predictive model, we have also transformed some data fields to binary (0, 1). The transformation will help us to determine all the required variables that impact the CT productivity, e.g., IsMount/IsLoad, IsRMG/IsRTG, the proportion of queuing out of total activities, the proportion of shuffling out of entire activities, etc. The final target variable is the unproductivity_flag, a binary value (0, 1). Each row in the dataset will present a job done by CT, which includes the source location, destination location, CT ID, start time, and end time of the job, as well as the other resources such as YC or QC involved. Each row will contain more than 40 columns of variables with a few hundred thousand records which represent a week of operations data from the container terminal. If the productivity is less than the cutoff value of 3.4 (a number given just for reference), the unproductive flag is 1, which means it is unproductive. Otherwise, if the productivity is more than or equal to 3.4, the unproductive flag is set to zero, which means it is a productive move. We are interested in building a predictive model to identify the unproductive movement and the factors affecting CT productivity.

If (productivity is < 3.4) then unproductivity_flag = 1, else unproductivity_flag = 0;

Once the data is loaded in the SAS Enterprise Miner software, we set the unproductivity_flag as TARGET and the rest of the variables as input. We also define the correct role for their respective data fields, such as binary, interval, or nominal variable. Data is being partitioned into 70% for training and 30% for validation.

The decision tree result shows that the significant factor influencing unproductivity_flag is 'Waiting Time' followed by 'Travel Time'. If we have a high value of traveling time and waiting time, then the unproductivity_flag will be 1, which means the truck is unproductive or has low productivity. We also observed a similar result in the initial data analysis and visualization, where we found out that high waiting time and travel time are the primary root cause for low productivity for each terminal.

Since the target variable unproductivity_flag is binary, Logistic Regression Model is used for prediction. We run the step-wise, forward, and backward logistic regression models and the forward regression model gave us the best results.

If we focus on the above-highlighted variables, we get more specific factors that reduce CT productivity. We can interpret these factors through the transformation output table, which has a value label for each variable. E.g., IsLoading – 0 means it is not a loading job. isRMG – 1 means that YC is RMG equipment.

The factors affecting CT productivity are as followed. If we drill down further on the type of yard activities variables, we get more specific factors that reduce CT productivity.

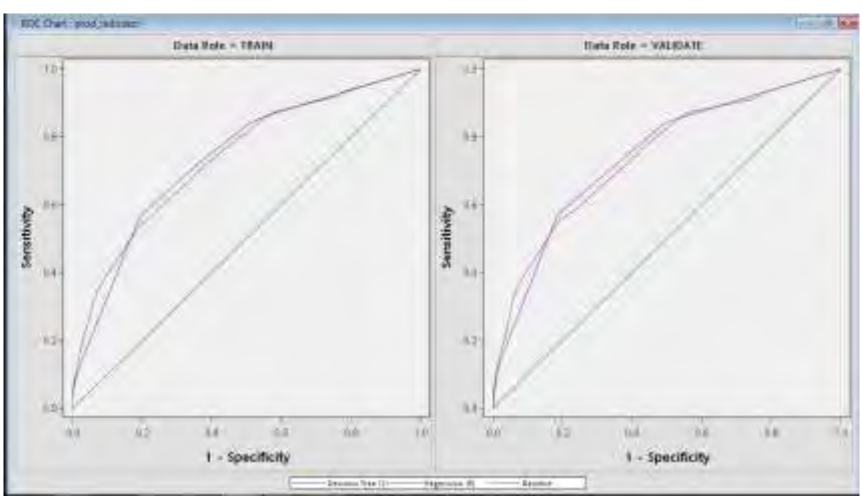
- Yard Crane is RMG
- Cross Gantry movement
- Logoff
- Mounting jobs
- Unplanned shuffling
- Queuing

Finally, we would like to compare the performance of two predictive models: LR and DT. Model comparison in SAS Enterprise Miner will be used, and the performance was based on the below parameters:

- ROC curve: It is a plot of the true positive rate against the false positive rate for the different possible cut points of a diagnostic test. It shows the trade-off between sensitivity and specificity. ROC curve gives the best-fit curve for each model. The better the curve, the better the model. To identify the best model, below are the two factors we observe in the ROC curve:
    - The closer the curve follows the left-hand border and then the top edge of the ROC space, the more accurate the test is.
    - The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test is.

The ROC curves show that Decision Tree has a slightly better curve toward the top than the Logistic Regression model.

Figure 5: ROC curves for model comparison



- **Misclassification rate:** Since the ROC curve was quite close for both models, we could not differentiate much about which is a better model. Thus, we will use another performance indicator for comparison, i.e., the misclassification rate of both models. The result showed that the misclassification rate for the forward logistic regression model (0.32473) is higher than the decision tree (0.3080). Hence, a decision tree is the champion model for this prediction. On the other hand, based on the misclassification rate of 0.30 for DT, the accuracy rate of the DT model is 1-0.3 = 0.7 (70%), which is considered quite suitable for the operation.

Table 2: Misclassification rate for model comparison					
Selected model	Model description	Valid: Average square error	Train: Average square error	Train: Misclassification Rate	Valid: Misclassification Rate
Y	Decision Tree	0.20191	0.20413	0.31109	0.30803
	Regression	0.20599	0.20731	0.32770	0.32473

5. CONCLUSION

In conclusion, to resolve the issue to increase CT productivity, we can follow a bottom-up solution approach. We already know that the significant problem for CT waiting time is a prolonged delay in YC activities. Therefore, we should focus on the basic YC activities, which leads to long CT waiting time. First, the terminal manager should take specific optimization measures to reduce the gantry movement and random shuffling, which are the main reasons for CT waiting time. Therefore, the below approach can be considered and worked upon to resolve the bottleneck at the yard.

Gantry Problem: Yard optimization can be implemented to determine the number of yard cranes required to minimize the CT waiting time to less than 15 minuters per trip. Thus, the outcome will lead to improved CT productivity.

Shifting/Shuffling Problem: This can be resolved by better planning the yard so that YC does not have to unnecessarily shuffle or shift among the containers for the mounting job. But, again, this is more of an operational solution that the operations team can consider.

Once these two issues are resolved, this will automatically lead to lesser vessel mounting jobs. In addition, reduced time in mounting jobs and more distributed jobs among various YCs will lead to more secondary activities performed per YC. As a result, less unproductive move will ultimately lead to a lower truck waiting time and will, in return, increase CT productivity.

In this paper, we focus on the impact of yard activities on CT productivity at various terminals. As part of yard activities, we identified that CT productivity is low because CT has to wait for the YC to serve it when YC is busy doing other jobs. Thus, more yard cranes are needed in each block during certain peak hours. However, YC are expensive resources, thus by finding the optimal number of YC required at each block in advance so that all activities are handled quickly, thus minimizing the waiting time for CT.

Furthermore, we can propose a dynamic allocation of YC to CT to reduce the waiting time. Finally, a discrete event simulation model can be developed to test the proposed model's performance. We can look at the queuing time of CT, system time as well as the queue length for each CT. The simulation will be subjected to various factors like block size, types of activities happening at each block, peak activity hours, weather conditions, etc., since these will impact the load on yard cranes at each time. Developing discrete event simulation models to test out the various operational scenarios and what-if analysis will be the direction for our future results.

REFERENCES

Chen, L. , Bostel, N. , Dejax, P. , Cai, J.Xi, L., A tabu search algorithm for the integrated scheduling problem of container handling systems in a maritime terminal, *European Journal of Operational Research*, 181 (1), pp 40-58, 2007.
Chung, R. K. , Li, C.-I.Lin, W., Interblock Crane Deployment in Container Terminals, *Transportation Science*, 36 (1), pp 79-93, 2002.
Das, S. K.Spasovic, L., Scheduling material handling vehicles in a container terminal, *Production Planning & Control*, 14 (7), pp 623-633, 2003.
Eleni Hadjiconstantinou & Nang Laik Ma, Evaluating straddle carrier deployment policies: a simulation study for the Piraeus container terminal, *Maritime Policy & Management*, 36(4), pp 353-366, 2009.
Gambardella, L. M. , Mastrolilli, M. , Rizzoli, A. E.Zaffalon, M., An optimization methodology for intermodal terminal management, *Journal of Intelligent Manufacturing*, 12 (5), pp 521-535, 2001.
Guan, Y.Cheung, R. K., The berth allocation problem: models and solution methods, *OR Spectrum*, 26 (1), pp 75-92, 2004.
Kim, K. H.Moon, K. C., Berth scheduling by simulated annealing, *Transportation Research Part B: Methodological*, 37 (6), pp 541-560, 2003.
Kim, K. H.Kim, K. Y., Optimal price schedules for storage of inbound containers, *Transportation Research Part B: Methodological*, 41 (8), pp 892 – 905, 2007.
Kim, K. H. , Kang, J. S.Ryu, K. R., A beam search algorithm for the load sequencing of outbound containers in port container terminals, *OR Spectrum*, 26 (1), pp 93 – 116, 2004.
Laik, N.Hadjiconstantinou, E., Container assignment and yard crane deployment in a container terminal: A case study, *Maritime Economics & Logistics*, 10, pp 90-107, 2008.
Laik, N., Hadjiconstantinou, E. Evaluation of operational plans in container terminal yards using Discrete-Event Simulation. *OR Insight*, 21, pp 10–18, 2008.
Stahlbock, R.Voß, S., Operation Research at Container Terminals - A Literature Update*, OR Spectrum*, 30 (1), pp 1-52, 2008.
Steenken, D. , Voß, S.Stahlbock, R., Container terminal operation and operations research - a classification and literature review, *OR Spectrum*, 26 (1), pp 5-49, 2004.
Vis, I. F. A.de Koster, R., Transshipment of containers at a container terminal: An overview, *European Journal of Operational Research*, 147 (1), pp 1-16, 2003.
Zhang, C., Liu, J. y., Wan, Y.-w. , Murty, K. G.Linn, R. J., Storage space Allocation in container terminals, *Transportation Research Part B: Methodological*, 37 (10), pp 883-903, 2003.

DECISION SCIENCES INSTITUTE
Analysis of Programming Developers' Information Needs through Web Mining

Ching-Chin Chern
National Taiwan University
Email: cchern@ntu.edu.tw

Yun-Wei Wu
National Taiwan University
Email: r09725059@ntu.edu.tw

Yu-Ru Wang
National Taiwan University
Email: r11725043@ntu.edu.tw

Chun-Wei Tsao
National Taiwan University
Email: r11725046@ntu.edu.tw

Bo Hsiao
National Yunlin University of Science and Technology
Email: bhsiao@yuntech.edu.tw

ABSTRACT

As developing issues are getting complicated, programming developers tend to seek experienced developers in the programming communities such as Stack Overflow for help. However, the forum's declining answer rate is making information retrieval more and more difficult. Thus, we aim to find developers' critical needs and the reasons for the declining answer rates to provide guidance for complementing related information. This research provides several methods and conclusions on developers' needs. We expect that the findings in this research can be adopted to improve developers' information needs, which results in a better working environment for developers.

KEYWORDS: Topic Modeling, Latent Dirichlet Allocation, Programming Community, Trend Analysis, Python

INTRODUCTION

Millions of issues could occur during the program development, but how do program developers seek help? If the program developers are familiar enough with the development tools, they might refer to the official documents written by the tools' developers. These documents include API documents, README files, and tutorials with sample codes. Some developers might search for related information on search engines which filter the information according to the keywords they input. However, these existing resources do not always fulfill all the information needs of developers. For example, the ambiguity and incompleteness of API documents may confuse users when reading them (Uddin and Robillard, 2015). In addition, search engines may fail when the desired information is not in their database because of the lack of updating or the users have needs in specific conditions. Whatever the reason is, these cases all show that existing resources cannot always satisfy program developers' needs. Thus, advice from a human is needed. Programming communities are the solutions in the way that they play vital roles as platforms for developers to exchange their opinions. Communities gather users who have similar social circles, interests, or professionals. This feature is essential for programming communities since developers care about the experiences and the professions of reviewers who give advice (Kovalenko et al., 2020). Moreover, prior work showed that developers seek information from documents and people (Hertzum and Pejtersen, 2000), which are both provided by these communities. Users can find experienced people to answer their questions. At the same time, all posts are stored on the website, thus are usable as documents. As programming communities become more popular, more developers join such platforms to seek help. According to Stack Overflow (SO) statistics, it attracts more than 100 million visitors per month. Furthermore, new questions are proposed every 13.6 seconds, showing the popularity of such communities (Stack Overflow, 2021).

However, not every question posted on websites is answered. According to past researches (Chua and Banerjee, 2015; Mondal et al., 2021; Saha et al., 2013; Shao and Yan, 2017), the number of unanswered questions on SO, the most popular programming forum globally, had gradually increased to 29% in 2017, and the number is still rising. Furthermore, the increasing number of developers with diverse backgrounds lead to various complicated demands, increasing the possibility of questions not being answered. According to a survey in 2020 from SO, 17% of respondents have coding experiences of less than five years, and nearly half of respondents started coding after 2010, which shows that plenty of people had joined in programming communities in the past ten years. Furthermore, nearly 40% of users major in neither computer science, computer engineering, nor software engineering, and the rate is slightly increasing from 2018 to 2020 (Stack Overflow Developer Survey, 2021). Since developers' needs are changing and new needs have appeared these years (Ali and Linstead, 2020; Panichella and Zaugg, 2020), the issues mentioned in the previous paragraphs will occur much more frequently.

We should better understand developers' needs, especially for those needs not being fulfilled. Moreover, we aim to find out the reasons of unfulfillment based on previous researches (Asaduzzaman et al., 2013; Baltadzhieva and Chrupała, 2015; Calefato et al., 2015; Calefato et al., 2018; Chua and Banerjee, 2015; Mondal et al., 2021; Saha et al., 2013; Shao and Yan, 2017). Thus, this research focuses on two key points: finding critical needs for program developers and the reasons for them not being fulfilled.

First, this research finds the critical needs for program developers. Since many studies have been dedicated to this field, we will primarily refer to the approaches from prior work in this step (Ali and Linstead, 2020; Allamanis and Sutton, 2013; Bajaj et al., 2014; Barua et al., 2014; Chauhan et al., 2019; Lee et al., 2018; Martinez and Lecomte, 2017; Mehrab et al., 2018; Shao and Yan, 2017; Shihab, 2015). We consider approaches in three aspects: data collection, data preprocessing, and model to find the most suitable combination for our research based on these three aspects. Then, we will verify the approaches' feasibilities and effectiveness by the evaluation of the models and the validation of their results. After extracting the main needs of program developers, we will analyze them according to their trends. Prior studies found that different topic development trends could reflect certain information such as rising and exhausted (Ali and Linstead, 2020; Bajaj et al., 2014, 2020). Thus, we will explore the topics with unique growing patterns and investigate the reasons to find out their needs.

Second, this research finds the reasons why needs are unfulfilled. We extend this problem based on the previous researches with two aspects (Asaduzzaman et al., 2013; Baltadzhieva and Chrupała, 2015; Calefato et al., 2015; Calefato et al., 2018; Chua and Banerjee, 2015; Mondal et al., 2021; Saha et al., 2013; Shao and Yan, 2017): the relationship between topics' features and their answer rates and the relationship between askers' abilities of presenting their needs and answer rates.

RESEARCH PROBLEMS AND HYPOTHESES

The two main aims of this research are finding the most critical needs of program developers and why some of them are not being fulfilled. We first find the critical needs using topic modeling and then compare them based on their essential features to find the reasons for unanswered questions. To find the most critical needs, we will first apply topic modeling and analysis on trends. The essential issue in topic modeling is to find the most relevant data and approach for our research. Next, we may extract the main topics in the discussions with the chosen data and approach. Moreover, we will conduct the validation for the model using the similarity of tags. After extracting the main topics, we aim to know the trend of each topic over time to see their developments. We are curious about the relationship between the answer rate and the trend of each topic. Besides the general development of topics, we especially care about the new topics with rising discussion trends since they represent the new information needs for developers. With the extracted critical needs using the approaches above, we will move to the other key point of this research: finding reasons for unanswered questions. We divide this part into two aspects: topic-related features and asker-related features.

There are so many posts in programming communities that new questions are posted every 13.6 seconds (Stack Overflow, 2021), which is even faster than the time that a person spends to read a post, not to mention to classify it. Thus, we can never catch up with the trends of discussions if we categorize the new questions manually. As a result, we should extract the topics from discussions by applying topic modeling approaches, which categorize contents automatically. Though many similar experiments have been conducted in the prior work, there is no standard operation process for dealing with this problem. Methods from previous studies are slightly different for different datasets and purposes. Thus, we need to find the most suitable topic modeling approach for our research. Based on previous research, we found issues to be dealt with in two aspects: data and approach. For data, there are two dimensions, including content and preprocessing method. For the approach, we need to solve issues including method selection, parameters, and evaluation methods. Since different types of data and approaches are suitable for different situations, we select the most fitted topic modeling approach via conducting experiments.

We first apply answered posts with an accepted answer as training data and then classify the rest unanswered with the trained model. Meanwhile, we verify the model using cosine similarity with tags distribution in each tag to ensure the model's effectiveness.

RQ1: Finding topics of developers' needs.

RQ2: Verifying the feasibility of applying a model trained with answered questions on unanswered questions.

After extracting the topics, we analyze the growing trend of each topic by focusing on three research questions: detecting the general pattern of development, identifying the new, popular topics, and determining the exhausted topics.

RQ3: Detecting the general pattern of development trend.

RQ4: Identifying new, popular topics.

RQ5: Determine exhausted topics.

We analyze what features would influence the answer rates of topics by studying how the features of topics influence each other. We apply difficulty, participation, and popularity as factors, listed in Table 1, and propose related research questions.

RQ6: Analyzing the features of topics.

Hypothesis 1: There is a negative relationship between the topic's difficulty (TD) and answer rates (TAR).

Hypothesis 2: There is a negative relationship between the topic's difficulty (TD) and the number of answers (TA).

Hypothesis 3: There is a positive relationship between the topic's answer rate (TAR) and the number of answers (TA).

Hypothesis 4: There is a positive relationship between the topics' difficulty (TD) and the number of views (TV).

Table 1: Topic Variables

Name	Symbol	Description
Difficulty	TD_t	TD_t denotes the difficulty level for a post to be answered in topic t.
Popularity	TV_t	TV_t denotes the number of views in topic t.
Participation	TA_t	TA_t denotes the number of answers in topic t.
Answer Rate	TAR_t	TAR_t denotes the answer rate in topic t.

Besides topic-related features, we are also curious about how the backgrounds of askers impact the answer rate. This section will present the features representing askers' experiences, which are question records, participation, and question quality, to see how they perform differently on answered and unanswered questions.

RQ7: Comparing askers' answered records between answered and unanswered questions.

Hypothesis 5: Answered questions have higher askers' answered rates (AAR) than unanswered questions.

Hypothesis 6: Answered questions have more askers' answered records (AA) than unanswered questions.

Hypothesis 7: Answered questions have fewer askers' unanswered records (AU) than unanswered questions.

Table 2: Askers' Background Variables

Category	Name	Symbol	Description	Reference
Participation	Asker's Proposed Question Records	AQ_a	AQ_a denotes the number of questions proposed by asker a.	(Asaduzzaman et al., 2013; Chua and Banerjee, 2015; Shao and Yan, 2017)
Question Records	Asker's Answered Records	AA_a	AA_a denotes the number of answered questions proposed by asker a.	(Asaduzzaman et al., 2013; Saha et al., 2013)
	Asker's Unanswered Records	AU_a	AU_a denotes the number of unanswered questions proposed by asker a.	(Saha et al., 2013)
	Asker's Answered Rate	AAR_a	AAR_a denotes the ratio of questions proposed by asker a.	(Mondal et al., 2021; Saha et al., 2013)
Question Quality	Asker's Tags Records	AT_a	AT_a denotes the number of tags added for a question proposed by asker a.	(Asaduzzaman et al., 2013; Chua and Banerjee, 2015; Shao and Yan, 2017)
	Asker's Questions Length	AL_a	AL_a denotes the number of terms in a question proposed by asker a.	(Asaduzzaman et al., 2013; Baltadzhieva and Chrupała, 2015; Calefato et al., 2015; Calefato et al., 2018; Chua and Banerjee, 2015; Saha et al., 2013; Shao and Yan, 2017)
	Asker's Uppercase Ratio	AUR_a	AUR_a denotes the ratio of uppercase letters in a question proposed by asker a.	(Calefato et al., 2015; Calefato et al., 2018)
	Asker's Code Snippets Rate	ACR_a	ACR_a denotes the ratio of questions containing code snippets proposed by asker a.	(Asaduzzaman et al., 2013; Baltadzhieva and Chrupała, 2015; Calefato et al., 2018; Chua and Banerjee, 2015; Mondal et al., 2021; Saha et al., 2013; Shao and Yan, 2017)

RQ8: Comparing askers' participations between answered and unanswered questions.

Hypothesis 8: Answered questions have more askers' asking records (AQ) than unanswered questions.

RQ9: Comparing askers' question quality between answered and unanswered questions.

Hypothesis 9: Answered questions have higher askers' code snippets rates (ACR) than unanswered questions.

Hypothesis 10: Answered questions have fewer askers' tags (AT) than unanswered questions.

Hypothesis 11: Answered questions have less askers' uppercase ratio (AUR) than unanswered questions.

Hypothesis 12: Answered questions have shorter askers' question lengths (AL) than unanswered questions.

RESEARCH METHODOLOGY

We choose Python-related posts on Stack Overflow for the data of our experiments. To ensure completeness of the topic development, we aim to the longest period possibly. Thus, we select the data from August, 2008, when SO was established, to June, 2021. Although we aim to gather the latest data to find out the current needs, posts might lack answers because they are newly asked. Previous research defines a post as unanswered if it remains unanswered for at least one month (Asaduzzaman et al., 2013). To be more conservative, we expand the period from one month to three months to guarantee that a post can be viewed as completed. Since the experiments of this research started in September, 2021, we thus collect the data until June, 2021, which is three months before the experiments began.

We choose the LDA model as the approach for topic modeling because we want to avoid issues that manually categorizing is too time-consuming and tag clustering is targeting more general, abstract concepts relying on users' expertise. LDA is a kind of unsupervised learning algorithms commonly used for extracting topics from the corpus. It assumes that each latent topic in all documents can be represented as a distribution over terms, and the term distributions of topics would also follow the same Dirichlet distribution (Jelodar et al., 2019). This model not only extracts topics from contents but also avoids human interference, the potential issues from the previous two methods.

We verify the trained model using failed questions with tags of the posts. Tags are standardized critical concepts of the question added by askers manually (Beyer and Pinzger, 2016; Gajduk et al., 2013; Treude et al., 2014, 2011). We do not directly use tags for categorizing topics because they usually represent broad concepts and lack details, compared to contexts. However, they can still be considered as the labels, or the key concepts, of the posts. Furthermore, they are added with human judgment, increasing their reliabilities.

We use the categorized results in the previous subsection to find the movement of each topic. The main variables we focus on are Impact Score and Answer Rate. We analyze the movements with line charts and the Cox-Stuart trend tests and then discuss those topics with the most significant changes. We follow the approaches used in Barua, Thomas, and Hassan's research (Barua et al., 2014), with small adjustment. We collect and summarize the monthly statistics for each topic. For example, statistics for January, 2000 are calculated with the questions asked from 1st to 31st January, 2000. The Impact Score for topic t in month m, $ImpactScore(t, m)$, is computed as the number of posts for topic t in month m divided by the number of posts in month m. The

Answer Rate for topic *t* in month *m*, *AnswerRate*(*t, m*), is computed as the number of successful posts for topic *t* in month *m* divided by the number of posts for topic *t* in month *m*. We then verify the increasing and decreasing trend movement of each topic using the Cox-Stuart trend test (Barua et al., 2014). We aim to classify the topics by their trends to see if the developing patterns differ for certain kinds of topics.

We select five topics with the largest and least $Diff_{ImpactScore}$ as the topics with the most significant growth and fall (Barua et al., 2014). The $Diff_{ImpactScore}(t)$ is computed as *ImpactScore*(*t*, 2021.06) minus *ImpactScore*(*t*, 2009.01). Since the data size significantly impacts the Impact Score for each topic, we aim to avoid the data with high variation in data size. Thus, we adopt data from January, 2009 as the first period to avoid the bias caused by the first few months after SO started operating in August, 2008.

It is reasonable to assume that the newest and most popular topics have increasing Impact Scores because they gather more discussions, while the exhausted topics may have decreasing ones. Despite trend tests and the calculation of differences, we take the topics' trend line graphs and their characteristics into account for their significant growth and falls as hypothesized in RQ4 and RQ5. The statistics using line charts by months for each topic is presented and the topics are compared by their Impact Scores to see if there are similarities or differences. Moreover, we find general developing patterns by observing the trends of the topics for hypothesis in RQ3.

The topics' variables, listed in Table 1, are *TD*, *TV*, *TA*, and *TAR*. When calculating for each observation, we need to find the statistics that can represent the topic. Note that the attributes *Difficulty*, *ViewCount*, and *AnswerCount* refer to attributes in Table 3. We calculate both mean and median for each topic and apply them for the following tests as usually did in prior studies to represent the performance of each topic (Ali and Linstead, 2020; Mehrab et al., 2018). Since the calculation of *TD* involves the duration of getting an accepted answer, we only take answered questions into account.

Table 3: Attributes of Topic Variables

Symbol	Definition
$TD_{t,\ mean}$	Mean of the *Difficulty* for questions in topic *t*
$TD_{t,\ median}$	Median of the *Difficulty* for questions in topic *t*
$TV_{t,\ mean}$	Mean of the *ViewCount* for questions in topic *t*
$TV_{t,\ median}$	Median of the *ViewCount* for questions in topic *t*
$TA_{t,\ mean}$	Mean of the *AnswerCount* for questions in topic *t*
$TA_{t,\ median}$	Median of the *AnswerCount* for questions in topic *t*
TAR_t	The number of answered questions in topic *t* divided by the number of all questions in topic *t*

After obtaining *TD*, *TV*, *TA*, and *TAR* for all topics, we test the relationship between attributes based on our assumptions in RQ6. Since our goals are to determine whether there exists a positive or negative relationship between the attributes, we will conduct tests of correlation coefficients. Before the tests, we first illustrate the scatter plots between the two variables to visualize the distributions and verify whether there are positive or negative linear relationships. In addition, except for *TAR*, we present the results using both mean and median.

After collecting the features above, we follow prior works, dividing all questions into two groups, answered and unanswered (Chua and Banerjee, 2015; Mondal et al., 2021; Saha et al., 2013). Then we compare the distribution between the two groups among all attributes to validate our hypotheses. Since all features would be interval data after processing, we need to check their normality further. If the data from both answered and unanswered groups follow normal distributions, we may apply a t-test to verify the means differences between the two groups. Otherwise, we will use non-parametric methods, Mann-Whitney U (MWU) Test with Cliff's Delta, to check if the distributions between two groups differ (Mondal et al., 2021). After deciding the method for comparing the two groups, we define the null and alternative hypotheses according to Hypothesis 5 to Hypothesis 12 from RQ7 to RQ9. We conduct the tests at a significance level of 0.05 and discuss the results of the tests compared to our assumptions.

EXPERIMENTS AND RESULTS

We collect Python-related posts asked on SO from August 2008 to June 2021 with 1,897,336 in total, of which 991,597 are answered and the rest, 905,739 posts unanswered (StackExchange Data Explore, 2022). Thus, the overall answer rate is 52.26%. We observe the posts' trends corresponding to the year, shown in Figure 1. The total number of posts kept increasing and reached 300k in 2020, showing the high popularity of Python and SO. As for the answered and unanswered posts, we found that the answered ones increase more than the unanswered ones. Thus, we may assume that the answer rate keeps declining. We confirm this assumption by the line chart of the annual answer rate, shown in Figure 2, and the result matches the finding of previous studies (Mondal et al., 2021; Saha et al., 2013; Shao and Yan, 2017).

However, it is worth to mention that we did not collect the full-year data in 2008 and 2021. Thus, the statistics in the two years are for reference and cannot represent their actual full-year performances. Moreover, we may observe that the answer rate dramatically dropped from 2020 to 2021.

We collect the asking records of askers who proposed Python-related questions on SO from January 2020 to June 2021 (StackExchange Data Explore, 2022). During the period, we collect 477,503 Python-related posts, of which 209,756 are answered and the other 267,746 are unanswered. Based on these samples, we further collect their askers' asking records within two years to represent their asking backgrounds. Consequently, other related 5,213,228 posts are collected.

The goal of model selection for topics extraction is to find the combination of low perplexity and high coherence score. First, we compare the performances of each dataset. There are four kinds of candidate datasets composed of different answer and code snippet adoptions. For consistency, the performances of each dataset concludes that the combination accepted answer with code snippets performs the best. Moreover, there is no big differences between each dataset's performance. Compared to perplexities, the values of coherences diverse more, and thus, we mainly compare the results with coherence scores. As for the selection of answers and code snippets, the performance of keeping code snippets is better than that of removing and adopting only accepted answers performs slightly better than applying all answers.

We name each topic manually for better discussion in the following sections. Since we verified the model with tags, besides representative terms, we also consider tags when naming. We found that some topics are mainly related to specific packages during the naming process and thus these topics are duly named (e.g., Tensorflow, PyQt, and tkinter).

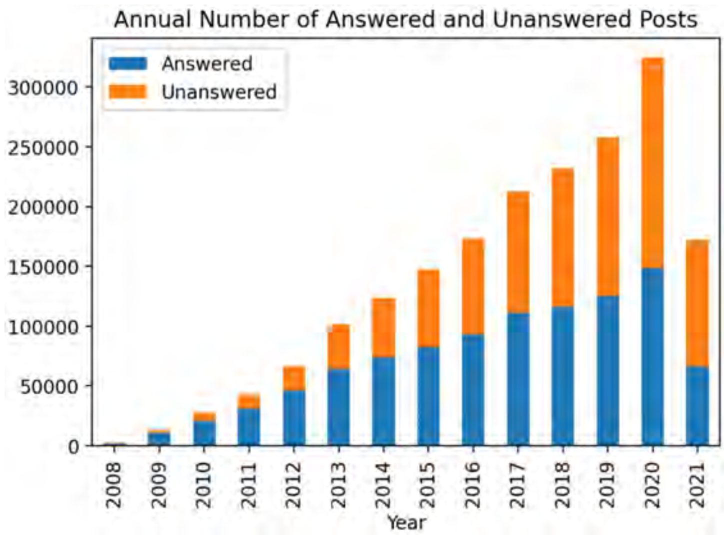


Figure 1: Annual Number of Answered and Unanswered Posts

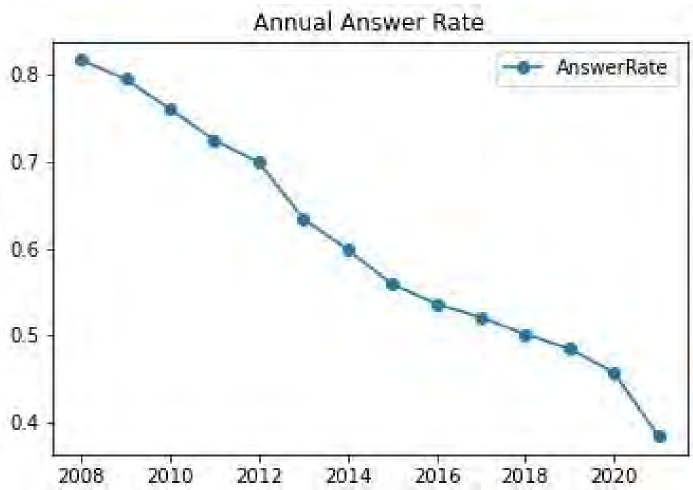


Figure 2: Annual Answer Rate

We categorize all posts into the forty topics and then calculate the monthly Impact Score and Answer Rate for each topic for the Cox-Stuart trend test. We list the topics that significantly increase or decrease in the Impact Score at a significance level = 0.05 in Table 4.

As for the Answer Rate, all topics are found to have significantly decreasing trends. That the Answer Rates of all topics are declining matches our finding in Figure 2. The result implies that the decline in overall Answer Rates is not because of some topics but rather, it is a widespread phenomenon among all topics

Next, we calculate the $Diff_{ImpactScore}$ for each topic. We list five topics with the most and the least differences in Table 5. We confirm that the topics filtered by $Diff_{ImpactScore}$ in Table 5 have the same developing trend direction as the results in Table 4. We further check their development trends using line charts to verify whether our conclusions for Table 8 are reasonable.

Table 4: Cox-Stuart Trend Test of Impact Score

Increasing				Decreasing			
Id	Name	Id	Name	Id	Name	Id	Name
2	Web-Scraping (HTML)	23	Datetime	1	Encode	24	Path and OS
4	Tensorflow	25	SQL	5	Log	26	Object-Oriented Programming
7	DataFrame	27	Turtle Graphics	11	Cython	28	Subprocess
8	Numpy Data Structure	30	Dictionary	12	String	29	Django (Request and Response)
9	tkinter	31	Geography	13	Decorator	33	Exception
10	Read File (Excel)	32	If-Else and For-Loop	17	General	36	Django
14	Image (Processing)	34	Cloud API	20	Web-Scraping (XML)	40	Python Installation
16	Login	35	Matplotlib				
19	Read File (CSV and txt)	37	OpenGL				
21	Module Error	38	NetworkX				
22	Multithreading						

Table 5: Most and Least DiffImpactScore

Increasing			Decreasing		
Id	Topic Name	$Diff_{ImpactScore}$	Id	Topic Name	$Diff_{ImpactScore}$
7	DataFrame	0.083568	17	General	-0.386032
4	Tensorflow	0.050795	13	Decorator	-0.023485
2	Web-Scraping (HTML)	0.039334	12	String	-0.013401
35	Matplotlib	0.037341	28	Subprocess	-0.012257
8	Numpy Data Structure	0.027603	24	Path and OS	-0.010318

Since each topic is viewed as an observation, all the posts are first categorized into the forty topics using the topic model trained and selected above and the variables in Table 3 of each topic is extracted. Note that TA_{median} is either 1 or 2, following the binary distribution. Moreover, for features TV and TD, their means are far higher than medians because these two features are significantly positively skewed. We plot the scatter plots of features involved in Hypotheses 1 to 4 to observe linear relationships between them and to decide how to conduct the coefficient correlation tests.

We further conduct the correlation coefficient tests for those pairs with linear relationships to verify our hypotheses. We list the alternative hypothesis of the further test more specifically in Table 6.

Since some hypotheses show linear relationships in more than one feature combination, we subdivide the four original hypotheses.

Table 6: Alternative Hypothesis of Topic Features

Hypothesis No.	Sub-hypothesis	Alternative Hypothesis
Hypothesis 1	1.1	$\rho(TD_{mean}, TAR) < 0$
	1.2	$\rho(TD_{median}, TAR) < 0$
Hypothesis 2		$\rho(TD_{median}, TA_{mean}) < 0$
Hypothesis 3		$\rho(TA_{mean}, TAR) > 0$
Hypothesis 4	4.1	$\rho(TD_{mean}, TV_{median}) > 0$
	4.2	$\rho(TD_{median}, TV_{median}) > 0$

We test whether the variables involved in Hypotheses 1.1 to 4.1 follow the normal distribution using the Shapiro test with a significance level = 0.05. We conclude that TAR and TV follow the normal distribution, while TA and TD do not. Therefore, we apply Spearman correlation test method for all the hypotheses because these attributes do not follow the normal distribution. The results are shown in Table 7.

Table 7: Results of Topic Feature Test

Hypothesis	Attribute 1	Attribute 2	Method	Correlation	p_value
1.1	TD_{mean}	TAR	Spearman	- 0.741	0
1.2	TD_{median}	TAR	Spearman	- 0.707	0
2	TD_{median}	TA_{mean}	Spearman	- 0.850	0.00005
3	TA_{mean}	TAR	Spearman	+ 0.597	0.00002
4.1	TD_{mean}	TV_{median}	Spearman	+ 0.873	0
4.2	TD_{median}	TV_{median}	Spearman	+ 0.816	0

Table 7 shows that all experiment results fit our assumptions to topic features. The conclusions from Hypotheses 1, 2, and 4 is that difficult topics (TD) have relatively lower answer rate (TAR), fewer answers (TA), and higher views (TV). It verifies our previous conjecture that difficulties could be a barrier to gathering answers. Moreover, their high popularities, represented by high TV, confirm the unsatisfied information needs of these topics.

Though the causal relationship cannot be confirmed, both situations show the insufficient information on the difficult topics. As for Hypothesis 3, the positive linear relationship between TAR and TA implies that the answer rate rises as the number of answers increases. It is a natural phenomenon since the more answers a post acquires, the more likely it obtains an accepted answer.

We divide samples into answered and unanswered groups and collect the asking records of the askers from these two groups. We calculate the basic statistics such as mean, median, and standard deviation and conduct the Shapiro normality tests for askers' background variables. Because none of these variables follow normal distribution, the Mann-Whitney U (MWU) test with

Cliff's Delta is applied to verify the hypotheses about the variables of askers' backgrounds that affect the answer rates.

The alternative hypotheses and the MWU test for each hypothesis with significance level at 0.05 with the Cliff's Delta are listed in Table 8. We conclude that all variables except AL match our assumptions. RQ7 considers the relationship between an asker's asking records and answered rate. The askers of answered posts have more answered records (AA), higher answered rate (AAR), and fewer unanswered records (AU) than the askers of unanswered posts. The Cliff's Deltas of AA and AAR are large, implying that these two variables are relevant when distinguishing the askers of answered posts and the askers of unanswered posts, while AU has relatively lower impact.

RQ8 focuses on the askers' participation on the platform. The participation of an asker in this research refers to AQ, or equivalently the frequency of an asker proposing questions. Via the MWU test, the askers of answered posts ask questions more frequently than the askers of unanswered ones. Thus, we conclude that these askers are more willing to interact with other users and more familiar with this platform. The open attitude and the familiarity with the platform's operation are the reasons for them to be more likely to receive answers.

Table 8: MWU Tests of Asker's Background Variables

Hypothesis No.	Alternative Hypothesis	MWU p_value	Cliff's Delta
Hypothesis 5	$P(AAR_{answered} > AAR_{unanswered}) > 0.5$	0	+ 0.851
Hypothesis 6	$P(AA_{answered} > AA_{unanswered}) > 0.5$	0	+ 0.524
Hypothesis 7	$P(AU_{answered} > AU_{unanswered}) > 0.5$	0	– 0.316
Hypothesis 8	$P(AQ_{answered} > AQ_{unanswered}) > 0.5$	0	+ 0.160
Hypothesis 9	$P(ACR_{answered} > ACR_{unanswered}) > 0.5$	0	+ 0.026
Hypothesis 10	$P(AT_{answered} > AT_{unanswered}) > 0.5$	0	– 0.012
Hypothesis 11	$P(AUR_{answered} > AUR_{unanswered}) > 0.5$	0	– 0.041
Hypothesis 12	$P(AL_{answered} > AL_{unanswered}) > 0.5$	1	+ 0.027

RQ9 analyzes the askers' asking styles. According to the results of the MWU test, the asking types of answered askers tend to contain code snippets (ACR), insert only relevant tags (AT), and avoid unnecessarily emphasized uppercase characters (AUR). However, the result of (AL) is inconsistent with our hypothesis and the claims in prior studies (Baltadzhieva and Chrupała, 2015; Calefato et al., 2018; Chua and Banerjee, 2015; Saha et al., 2013). The experiment result concludes that a longer question is more likely to be answered, as opposite from our hypothesis. As result, a more detailed description of the question is needed to gather other users' attention and understanding of the question due to the complicated usage of Python as time goes by. Therefore, a long question with more explanation may attract more answerers.

CONCLUSIONS

In this study, we study the critical needs of program developers and the reasons for their needs not being fulfilled. The main process includes (1) extracting critical topics using the LDA model and data collected from SO, (2) detecting the trend of each topic to find specific movements, (3) analyzing the relationship between topics' features, and (4) finding the askers' backgrounds leading to high answered rate.

This study analyzes the Python-related posts on Stack Overflow to find the main topics on the discussion as users' critical needs. We trained models with various datasets and parameters and select the one with the best performance. We conclude that data applying only accepted answers and containing code snippets and parameters composed of asymmetric α and symmetric β performs the best. We verify the model's feasibility using topics' tag distribution and prove that the model trained with answered posts categorizes the unanswered posts well.

With the Cox-Stuart trend test and line charts on $Diff_{ImpactScore}$ metric, we detect the new, popular topics, which are mostly related to Data Analytics, and the exhausting topics, which developed early and have matured with similar contents over time. The complex topics attract more views while having lower answer rates. Moreover, topics with low answer rates are likely to receive fewer answers. The phenomenon infers serious insufficient information for those difficult topics, providing a reference for the urgent level of topics' information needs.

Finally, evidence shows that askers with higher answered rates, more answered records, and better-asking habits are more likely to be answered. In addition, askers' previous answered rate most significantly impacts the answer rate of a new question. As a result, it is vital to assist askers with low answered rates to improve their asking abilities. Since the presenting styles of a post influence the possibility of receiving an answer, providing guidelines for users to ask questions can help those with low answered rates.

ACKNOWLEDGEMENTS

This research was sponsored by the National Science and Technology Council in Taiwan, under project number MOST 111-2410-H-002-067-MY2.

REFERENCES

Ali, R. H. and Linstead, E., "Modeling topic exhaustion for programming languages on StackOverflow," Proceedings of the 32nd International Conference on Software Engineering and Knowledge Engineering – SEKE'20, pp. 400-405, 2020.

Allamanis, M. and Sutton, C., "Why, when, and what: analyzing Stack Overflow questions by topic, type, and code," Proceedings of the 10th Working Conference on Mining Software Repositories - MSR '13, pp. 53-56, 2013.

Asaduzzaman, M., Mashiya, A. S., Roy, C. K. and Schneider K. A., "Answering questions about unanswered questions of Stack Overflow," Proceedings of the 10th Working Conference on Mining Software Repositories - MSR '13, pp. 97-100, 2013.

Bajaj, K., Pattabiraman, K. and Mesbah, A., "Mining questions asked by web developers," Proceedings of the 11th Working Conference on Mining Software Repositories - MSR '14, pp. 112-121, 2014.

Baltadzhieva, A and Chrupała, G., "Predicting the quality of questions on Stackoverflow," Proceedings of Recent Advances in Natural Language Processing, pp. 32-40, 2015.

Barua, A., Thomas, S. W. and Hassan, A. E., "What are developers talking about? An analysis of topics and trends in Stack Overflow," Empirical Software Engineering, Vol. 19, Issue 3, pp. 619–654, 2014.

Beyer, B. and Pinzger, M., "Grouping Android tag synonyms on Stack Overflow," Proceedings of the 13th IEEE/ACM Working Conference on Mining Software Repositories - MSR '16, pp. 430-440, 2016.

Calefato, F., Lanubile, F., Marasciulo, M. C. and Novielli, N., "Mining successful answers in Stack Overflow," Proceedings of the 12th Working Conference on Mining Software Repositories - MSR '15, pp. 430-433, 2015.

Calefatoa, F., Lanubileb, F. and Novielli, N., "How to ask for technical help? Evidence-based guidelines for writing questions on Stack Overflow," Information and Software Technology, 94, pp. 186-207, 2018.

Chauhan, A., Verma, A., Deepanshi and Gurve, M., "Tag based answer recommendation system," Proceedings of the 2019 International Conference on Signal Processing and Communication - ICSC '19, pp. 195-201, 2019.

Chua, A. Y. K. and Banerjee, S., "Answers or no answers: Studying question answerability in Stack Overflow," Journal of Information Science, Vol. 41, Issue 5, pp. 720-731, 2015.
Gajduk, A., Madjarov, G. and Gjorgjevikj, D., "Intelligent tag grouping by using an agglomerative clustering algorithm," Proceedings of the 10th Conference for Informatics and Information Technology - CIIT '13, pp. 94-96, 2013.

Hertzum, M. and Pejtersen, A. M., "The information-seeking practices of engineers: searching for documents as well as for people," Information Processing and Management, Vol. 36, Issue 5, pp. 761-778, 2000.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li1, Y. and Zhao, L., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," Multimedia Tools and Applications, Vol. 78, Issue 11, pp. 15169-15211, 2019.

Kovalenko, V., Tintarev, N., Pasynkov, E., Bird, C. and Bacchelli, A., "Does reviewer recommendation help developers?" IEEE Transactions on Software Engineering, Vol. 46, Issue 7, pp. 710-731, 2020.

Lee, M., Jeon, S. and Song, M., "Understanding user's interests in NoSQL Databases in Stack Overflow," Proceedings of the 7th International Conference on Emerging Databases, Springer, pp. 128-137, 2018.

Martinez, M. and Lecomte, S., "Discovering discussion topics about development of cross-platform mobile applications using a cross-compiler development framework," 2017.

Mehrab, Z., Yousuf, R. B., Tahmid, I. A. and Shahriyar, R., "Mining developer questions about major web frameworks," Proceedings of the 14th International Conference on Web Information Systems and Technologies, pp. 191-198, 2018.

Mondal, S., Saifullah, C. M. K., Bhattacharjee, A., Rahman M. M. and Roy, C. K., "Early detection and guidelines to improve unanswered questions on Stack Overflow," Proceedings of the14th Innovations in Software Engineering Conference - ISEC '21, pp. 1-11, 2021.

Panichella, S. and Zaugg, N., "An empirical investigation of relevant changes and automation needs in modern code review," Empirical Software Engineering, Vol. 25, Issue 6, pp. 1-40, 2020.

Saha, R. K., Saha, A. K. and Perry, D. E., "Toward understanding the causes of unanswered questions in software information sites- A case study of Stack Overflow," Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering - ESEC/FSE '13, ACM, pp. 663-666, 2013.

Shao, B. and Yan, J., "Recommending answerers for Stack Overflow with LDA model," Proceedings of the 12th Chinese Conference on ComputerSupported Cooperative Work and Social Computing - ChineseCSCW '17, ACM, pp. 80-86, 2017.

Shihab, E., "What are mobile developers asking about? A large scale study using stack overflow," Empirical Software Engineering, Vol. 21, Issue 3, pp. 1192-1223, 2015.

Treude, C., Barzilay, O. and Storey, M., "How do programmers ask and answer questions on the web? (NIER track)," Proceedings of the 33rd International Conference on Software Engineering - ICSE '11, pp. 804-807, 2011.

Uddin, G. and Robillard, M. P., "How API documentation fails," IEEE Software, Vol. 32, Issue 4, pp. 68-75, 2015.

Stack Overflow, accessed 2021-10-20. https://stackoverflow.com/.

Stack Overflow Developer Survey, accessed 2021-12-07.

StackExchange Data Explore, accessed 2022-03-29. https://data.stackexchange.com/stackoverflow/.

DECISION SCIENCES INSTITUTE

Planning for Donor Advised Funds: How, When, and Why

Nathan Oestreich
San Diego State University
drno@sdsu.edu

Sheldon R. Smith
Utah Valley University
smithsh@uvu.edu

ABSTRACT

Generous people have long enjoyed U.S. tax benefits from charitable giving. The simplest source of these activities always came from the itemized deduction for cash and property contributions, but not everyone has appreciated assets. Many other sophisticated devices exist. This research involves an investigation of donor advised funds that might be used to bunch deductions for people who might not itemize every year before and after recent tax reform.

KEYWORDS: Donor Advised Funds, Charitable Deductions, Time Value of Money, Bunching Deductions, Tax Reform

INTRODUCTION

A donor-advised fund, or DAF, allows a grantor to deposit assets with a sponsor organization and make direct donations to qualified charities (i.e., Section 501(c)(3) organizations) over time when the donor so chooses. The donor gets a tax deduction when the DAF is funded which can provide various benefits like bunching deductions (itemize in one year and claim the standard deduction in one or more alternate years. The sponsoring organization manages the account, and legally holds title to the assets, while the donor can advise on how to invest the assets and determines when and where to donate them (IRC § 4966) (Nerdwallet)

> Donor-advised funds are subject to new requirements under the Pension Protection Act of 2006. The IRS has issued guidance and new procedures implementing the legislation.
>
> - New excise taxes may apply to sponsoring organizations and managers of donor-advised funds.
> - Thus, transactions between sponsoring organizations and fund managers may be subject to intermediate sanctions excise taxes and donor-advised funds may be subject to taxes on excess business holdings.
> - Donors are provided guidance on how to determine whether a grantee that is a donor-advised fund is a public charity under section 509(a)(1), (2), or (3), in IRS Business Master File information.
> - Notice 2006-109 provides interim guidance on issues affecting supporting organizations and sponsoring organizations of donor advised funds.
> - Responding to Congressional direction, Treasury Department reported the results of a study of donor-advised funds and supporting organizations in December 2011." (Nerdwallet).

STEP BY STEP GUIDELINES

Nerdwallet provides these step-by-step guidelines as to how to select and invest in a donor-advised fund (Nerdwallet).:

Step 1. Compare donor-advised fund sponsoring organizations.

There are many different kinds of sponsoring organizations. Commercial donor-advised funds, for example, are run by nonprofit arms of national financial-services firms. We highlight three of them in the table below. As it shows, donor-advised funds make money from fees.

	Fidelity Charitable	Schwab Charitable	Vanguard Charitable
Minimum initial contribution	$0.	$0 for Core accounts; $250,000 for professionally managed accounts.	$25,000.
Minimum for additional contributions	$0.	$0.	$5,000.
Minimum grant to charity	$50.	$50.	$500.
Annual admin fee	Greater of 0.60% or $100 (tiered after $500,000).	0.60% (tiered after $500,000).	0.60% (tiered after $500,000).
Investment fees	0.015% to 0.99%.	0.03% to 0.77%.	0.03%.
Maintenance fee	$0.	$0.	$250/year if below $25,000.

Step 2. Contribute cash or other assets to the donor-advised fund.

You can put in cash, stocks or other investments, such as cryptocurrency or even your ownership in a private business. Note: Contributions are irrevocable, meaning that once you contribute the assets, you can't get them out again.

Step 3. Itemize your taxes to get the tax break.

That means filling out Schedule A when you do your taxes and making sure that your itemized deductions exceed the standard deduction to get the most bang for your donated bucks. (Learn more about how to decide whether it's worth it to itemize your taxes this year.) You receive your tax break in the year you contribute to your donor-advised fund. In case you're curious about limitations on deductions, you can deduct up to 60% of your adjusted gross income for cash contributions. If you're contributing securities or appreciating assets to your donor-advised fund, you can deduct up to 30% of your AGI. If you aren't able to get your full deduction in a single year, you typically have five years to claim the unused deductions.

Step 4. Help your donation grow.

Although the sponsoring organization controls the money in the donor-advised fund and the investment options you choose from, you get to recommend which investments to use. The assets in a donor-advised fund are then invested and can appreciate tax-free until you're ready to donate them to your charity of choice. Unlike private funds, there is no mandatory distribution date for individual donors, meaning the funds could sit in the donor-advised fund for years before charities receive them. However, most providers, such as Vanguard, Schwab and Fidelity, have policies that require you to disburse funds to charity regularly.
What are the advantages of waiting to disburse funds to charities? Some people prefer to wait for the investments to mature so they can give a larger amount. Others want the tax deduction immediately but need time to select charities they want to give to.

Step 5. Pick charities to support.

That's the goal of the entire process. You can support pretty much any IRS-qualified public charity. Typically, the entity that sponsors the donor-advised fund is responsible for checking out charities to ensure that the money goes to legit ones.

A cursory search of other sponsors located one, The San Diego Foundation, that does not openly disclose their minimum DAF contribution (San Diego) and Thrivent Financial that once had a minimum contribution, but now has none (Thrivent).

TAX REFORM

The benefits of DAFs have been impacted for many donors by recent tax law changes, especially the 2017 tax reforms.

The Tax Cuts and Jobs Act of 2017 (TCJA) is known for a large number of changes including temporary (generally 2018 – 2025) reductions in tax rates, repeal of the personal and dependency exemptions, increase in the child tax credit, and a substantial increase in the standard deduction. The other changes included in the bill, though massive, do not generally impact charitable deductions. Perhaps the most significant aspect is whether taxpayers itemize their deductions or not. Fewer taxpayers are itemizing their deductions after tax reform due to the enhanced standard deduction and the $10,000 limit on deductions for state and local taxes. This may be especially true for couples, as the $10,000 limit is the same for them as for singles and heads of household. These individuals may not have enough deductions to itemize or may only benefit from a fraction of their charitable contributions. However, if they take advantage of DAFs, they can get a tax benefit from all their incremental contributions by strategically itemizing in select tax years.

The increase in the number of taxpayers claiming the standard deduction, rather than itemizing, is revealed in the Internal Revenue Service's Statistics of Income Report.

Table 1. Individual Tax Return Statistics (selected).
(SOI, 2019, Individual Income Tax Returns Complete Report, Table A)

	2017	2018	2019
Number of individual tax returns filed	152,903,231	153,774,296	157,796,807
Number of individual tax returns claiming the standard deduction	104,013,115	134,271,137	138,307,604
Percentage of tax returns claiming the standard deduction	68.03%	87.32%	87.65%

PLANNING

Strategic planning should be a part of an over financial plan, including estate plans. For many generous donors who do not have appreciated assets and have not reached age 70.5, DAFs may provide significant benefits in those plans. Some of the factors are discussed here.

Itemizing in alternate years.

A generous standard deduction, compared to prior years, applies from 2018 - 2025. For 2023, the amount is $13,850 for single taxpayers. Singles who have itemized deductions up to that amount gain no benefit from itemizing. What to do?

Substantial DAF funding in year 1 and itemizing, followed by claiming the standard deduction in later years can result in substantial tax savings. Some commentators refer to this as "bunching."

Example

Wally, a single taxpayer, contributes significant amounts, say $6,000 per year, to a church. Wally could claim itemized deduction, but the total is less than the $13,850 standard deduction. If able, Wally could contribute $30,000 to a DAF and distribute $6,000 per year to the church. Wally could itemize in 2023 and claim the standard deduction in 2024 – 2027. The earnings on the DAF escape income tax.

This spreadsheet assumes 4% growth in the standard deduction, other itemized deductions, and charitable contributions. Implicit is an assumption that earnings in the DAF are 4% and those are given to charity.

Without DAF	2023	2024	2025	2026	2027
Standard Deduction	$ 13,850	$ 14,404	$ 14,980	$ 15,579	$ 16,203
Other Itemized deductions	$ 8,000	$ 8,320	$ 8,653	$ 8,999	$ 9,359
Charitable contributions	$ 6,000	$ 6,240	$ 6,490	$ 6,749	$ 7,019
Change in taxable income	$ 150	$ 156	$ 162	$ 169	$ 175
Related change in tax	$ 48	$ 50	$ 52	$ 54	$ 56

With DAF	2023	2024	2025	2026	2027
Standard Deduction	$ 13,850	$ 14,404	$ 14,980	$ 15,579	$ 16,203
Other Itemized deductions	$ 8,000	$ 8,320	$ 8,653	$ 8,999	$ 9,359
Charitable contributions	$ 30,000	$ -	$ -	$ -	$ -
Change in taxable income	$ 24,150	$ -	$ -	$ -	$ -
Related change in tax	$ 7,728	$ -	$ -	$ -	$ -

The net savings based on these assumptions are $7,468. Certainly, one should consider the time value of these cash flows.

Borrowing to fund. Time value of money.

Taxpayers may not have money laying around, and even if they do, the opportunity costs of using the money should be considered. Let's assume that the funds are borrowed at 6% interest. First of all, the interest would normally not be deductible.

With DAF	2023	2024	2025	2026	2027
Loan Balance	na	$ 24,000	$ 18,000	$ 12,000	$ 6,000
Interest at 6%	na	$ 1,440	$ 1,080	$ 720	$ 360

The total interest in this scenario is $3,600. Taking that amount from the net savings above still leaves $3,868 ($7,468 - $3,600). This represents a worst-case scenario and the net savings is more likely to be larger with careful planning.

QCDs for older taxpayers

Qualified charitable distributions (QCDs) are a factor for taxpayers approaching age 70.5 when they can make these tax-free contributions directly to charity. This is another option for nonitemizers who have traditional IRAs. Even if they have no IRAs, they can roll funds over to IRAs from other defined contribution plans (§§ 401(k), 403(b), or 457). However, they cannot use a QCD to fund a DAF.

Funding with appreciated capital assets.

The DAF can be funded with cash and other assets like appreciated investments. It cannot be funded with QCDs, but direct QCDs would be somewhat equivalent.

The deduction for contributions of most long-term capital assets is generally fair market value. The limitations on these contribution deductions are percentages based on adjusted gross income and are deemed to beyond the scope of this research at this time.

Future research will consider the effects of state and local income taxes. California, for example, does not have a preferred tax rate for capital gains. As a result, the DAF may provide exaggerated benefits beyond federal taxes. Taxpayers may avoid only 15% in federal taxes, but 6 – 10% in state taxes.

TAX UNCERTAINTY

The individual tax changes made by TCJA are currently scheduled expire after 2025, a mere two years away. No one knows whether the enhanced standard deductions, the limit on state and local tax deductions, or other changes will be extended. Will they be replaced with more moderate provisions? The authors do not predict tax law changes, but taxpayers are faced with any possibilities as they strategize in what are in effect multiple year scenarios.

ANONYMITY

How often do we receive solicitations from charities due to prior gifts or referrals? These endless requests might be limited by having DAF disbursements made anonymously.

CONCLUSION

The tax savings from DAFs for taxpayers who would not itemize or who have modest itemized deductions in excess of the standard deduction are not trivial. The consequences are greatly exaggerated by tax reform, particularly the enhanced standard deductions from 2018 to 2025. Planning is tenuous for 2026 which are scheduled end the 2017 individual income tax changes.

REFERENCES

Internal Revenue Code §§ 4966(d)(2) – (4)—See Appendix.

Nerdwallet (https://www.nerdwallet.com/article/taxes/donor-advised-funds).

The San Diego Foundation (https://www.sdfoundation.org/news-events/sdf-news/what-to-know-about-donor-advised-funds-in-2022-rules-tax-deductions-comparisons-and-more/#:~:text=Donating%20cash%2C%20via%20check%20or,have%20to%20grant%20to%20charity).

TCJA. Tax Cuts and Jobs Act of 2017 (TCJA)

Thrivent Financial [https://www.thrivent.com/insights/generosity-giving/donor-advised-funds-helping-you-help-others].

APPENDIX

IRC Section 4966 (d)(2) Donor advised fund

(A) In general. Except as provided in subparagraph (B) or (C), the term "donor advised fund" means a fund or account—

- **(i)** which is separately identified by reference to contributions of a donor or donors,
- **(ii)** which is owned and controlled by a sponsoring organization, and
- **(iii)** with respect to which a donor (or any person appointed or designated by such donor) has, or reasonably expects to have, advisory privileges with respect to the distribution or investment of amounts held in such fund or account by reason of the donor's status as a donor.

(B) Exceptions. The term "donor advised fund" shall not include any fund or account—

- **(i)** which makes distributions only to a single identified organization or governmental entity, or
- **(ii)** with respect to which a person described in subparagraph (A)(iii) advises as to which individuals receive grants for travel, study, or other similar purposes, if—
  - (C) such person's advisory privileges are performed exclusively by such person in the person's capacity as a member of a committee all of the members of which are appointed by the sponsoring organization,
  - (II) no combination of persons described in subparagraph (A)(iii) (or persons related to such persons) control, directly or indirectly, such committee, and
  - (III) all grants from such fund or account are awarded on an objective and nondiscriminatory basis pursuant to a procedure approved in advance by the board of directors of the sponsoring organization, and such procedure is designed to ensure that all such grants meet the requirements of paragraph (1), (2), or (3) of section 4945(g).

(C) Secretarial authority. The Secretary may exempt a fund or account not described in subparagraph (B) from treatment as a donor advised fund—

- **(i)** if such fund or account is advised by a committee not directly or indirectly controlled by the donor or any person appointed or designated by the donor for the purpose of advising with respect to distributions from such fund (and any related parties), or
- **(ii)** if such fund benefits a single identified charitable purpose.

IRC § 4966(d)(3) Fund manager. The term "fund manager" means, with respect to any sponsoring organization—
(A) an officer, director, or trustee of such sponsoring organization (or an individual having powers or responsibilities similar to those of officers, directors, or trustees of the sponsoring organization), and
(B) with respect to any act (or failure to act), the employees of the sponsoring organizationhaving authority or responsibility with respect to such act (or failure to act).

IRC § 4966(d)(4). Disqualified supporting organization.
(A) In general. The term "disqualified supporting organization" means, with respect to any distribution—
(i) any type III supporting organization (as defined in section 4943(f)(5)(A)) which is not a functionally integrated type III supporting organization (as defined in section 4943(f)(5)(B)), and
(ii) any organization which is described in subparagraph (B) or (C) if—
(C) the donor or any person designated by the donor for the purpose of advising with respect to distributions from a donor advised fund (and any related parties) directly or indirectly controls a supported organization (as defined in section 509(f)(3)) of such organization, or
(II) the Secretary determines by regulations that a distribution to such organization otherwise is inappropriate.
(B) Type I and type II supporting organizations. An organization is described in this subparagraph if the organization meets the requirements of subparagraphs (A) and (C) of section 509(a)(3) and is—
(i) operated, supervised, or controlled by one or more organizations described in paragraph (1) or (2) of section 509(a), or
(ii) supervised or controlled in connection with one or more such organizations.
(C) Functionally integrated type III supporting organizations**.** An organization is described in this subparagraph if the organization is a functionally integrated type III supporting organization (as defined under section 4943(f)(5)(B)).

DECISION SCIENCES INSTITUTE

Analyzing the effect of Avatar's social presence in customer online experiences and the Avatar usage intent in purchase intention with the mediating role of an entertaining and informative site

Ali Sanayei
Professor, University of Isfahan, Isfahan, Iran
Visiting Professor, University of Illinois, Chicago, USA
Email: a_sanayei@ase.ui.ac.ir

Behnoosh Ansari
M.A of E-Commerce, Hasht Behesht Higher Education Institute, Isfahan, Iran
Member of Information Technology Management Research Group, University of Isfahan
Email: Behnoosh_ansari@yahoo.com

ABSTRACT

One of the challenges of online sales is the absence of a person and the lack of social interaction. The solution to this problem is the use of avatars. Nowadays, the use of avatars on websites and social networks has increased. This research aims to investigate the impact of avatars on customers' online experience and purchasing intention. This research is applied in terms of its objective and descriptive survey regarding the data collection method. Furthermore, in terms of examining the relationships between variables, it is correlational. The study population includes individuals familiar with avatars and those who have had online shopping experiences through avatars. Due to the unavailability of a specific list of these individuals, non-random sampling was performed, and the sample size was 370 individuals. A researcher-made questionnaire was used to collect the data. The collected data were first analyzed using descriptive statistics and then inferential statistics using SPSS and AMOS software. Structural equation modeling was used to test the hypotheses. The results of hypothesis testing showed that the presence of avatars on websites and virtual networks with the mediating role of information and entertainment has a significant positive impact on customers' online experience. Additionally, the intention to use avatars also influences customers' purchasing intention.

KEYWORDS: Avatar, Online Experience, Intention to use Avatars, Purchasing Intention

INTRODUCTION

The increasing development of information technology in the e-commerce space provides significant potential for enhancing interaction between customers and organizations in creating value. One solution for increasing interaction is the use of avatars. An avatar is a graphical representation that can lead to increased satisfaction, positive attitude towards a product, and purchase intention. In recent years, customer behavior has changed in line with technological advancements and the global economic environment. Customers easily acquire specialized information through the internet and become knowledgeable about products. Consequently, they increase their pre-purchase awareness. This is why customers are not easily influenced by advertisements. In the past few years, especially after the COVID-19 pandemic, the use of the internet in commerce and trade has experienced a remarkable increase. In other words, the internet has created a new season in marketing. Companies can attract new customers through the use of the internet. Furthermore, by analyzing their online behavior and performance, companies can personalize communications, products, services, and prices (Dadvand et al., 2021). Traditional marketing methods are not feasible in this platform. Therefore, innovative marketing approaches such as avatar marketing have received significant attention.

Considering that the global internet network and websites provide businesses with opportunities to establish connections with customers, the level of interaction between websites and customers is among the important and key issues that companies face. Based on conducted research, a company's website will be successful when it has a high level of interaction with its customers. In simpler terms, the website should have created valuable and relevant content that meets the customer's needs, thereby building customer trust. Subsequently, this created trust can lead to a change in customer behavior and a decision to purchase a product or service from the website. The presence of an avatar as a salesperson may increase consumers' perception of social interaction in online purchasing and elicit their social responses or behaviors towards the website (Wang et al., 2007). The literature on consumer-seller communication indicates that consumers are interested in establishing social interactions with sellers and also reap social benefits from them. These social benefits include obtaining information and entertainment, which leads to enjoying their time with the seller. When consumers experience these social benefits, they develop positive emotions towards the store.

Avatars are used in online purchasing as sales assistants and representatives of the company, enhancing social interaction in online shopping. The presence of avatars in online purchasing can improve the shopping experience and facilitate social communication between customers and sellers. The use of avatars in the online domain helps build trust and increase user engagement. Additionally, avatars can enhance the customer experience by humanizing online stores and influencing customer behavior. Adopting a human-centric approach and incorporating elements such as emotive descriptions and relevant images of individuals, videos, and appropriate visuals related to avatars, enhances the online customer experience, ultimately leading to increased sales.

The increasing use of virtual avatars in the digital world raises the question of which avatars consumers prefer and trust. How do consumers develop trust and inclination to purchase products or services? Successful customer experience and willingness to make repeat purchases are essential aspects of contemporary marketing and can be considered a competitive advantage for the desired brand. Influencer marketing is based on real individuals and is executed by them. Therefore, there is limited information available regarding the factors that have contributed to the success of virtual influencers in advertising. This study aims to analyze the impact of avatar presence on customers' online experience and their inclination to use avatars and make purchases, with the mediating role of entertainment and information provision. While there is considerable research on psychological factors and marketing strategies related to human influencers, there is limited information available regarding virtual influencers. This research focuses on analyzing the influence of avatars on online retail and the inclination to use virtual influencers in the role of information providers and influencers in shaping purchase intent.

THEORETICAL BACKGROUND AND HYPOTHESES

Avatar marketing has gained momentum as a brand support strategy, but there is a lack of studies on its benefits, risks, and operational mechanisms. Silva et al. (2022) conducted a research study titled "Avatar Marketing: An Investigation of Interaction and Credibility of Virtual Influencers on Instagram" to understand the strategies employed by human-like avatars in influencer marketing. They collected data from the HypeAuditor platform and used descriptive statistics and semiotic analysis for analysis. The results showed that establishing a connection between avatars and followers relies on credibility attribution and alignment of posts with followers' lifestyles. Avatars serve as effective endorsers, utilizing strategies from humanization to robotization/meta-verses. Park et al. (2021) investigated the influence of avatar presentation, market status, and representative identity on perceived service quality and satisfaction with online chat services.

They found that when customers receive services from an artificial intelligence agent, they perceive better service quality but exhibit lower satisfaction. This pattern was observed only for companies with a good market status. The study aimed to differentiate customer chat services through interactive communication and new digital technologies. Molina (2021) explored the history and criteria influencing avatar perception in his study titled "Famous Avatars: A Technical Approach to Creating Digital Avatars for Social Marketing Strategies." He discussed avatar design, skills, personality traits, and their impact on social growth, persuasion, communication, and social interaction. The study highlighted how avatars have been utilized by various brands to demonstrate values, attract customers, and appear trustworthy. Moustakas et al. (2020) focused on virtual influencers, human-like avatars with significant followings on social networks. They examined the effectiveness of virtual influencer marketing through semi-structured interviews with experts in digital media. The study identified the potential advantages and disadvantages of using avatars as a marketing strategy. They concluded that Generation Y and Z are the most attractive target audience for avatars, and industries such as fashion, entertainment, lifestyle, healthcare, wellness, and tourism could benefit the most from using virtual influencers. Kravchik and Newhall (2020) discussed an avatar-based integrated marketing campaign targeting Generation Z, the current high school and college generation. The campaign aimed to attract and encourage Generation Z to study abroad by using avatars. The University of Minnesota implemented this campaign, introducing multiple avatars in print advertisements, informational materials, and social media platforms. The study presented the various methods used to develop the avatar-based campaign. Kumar (2020) examined how avatar characteristics and presence influence customer behavior during online interactions, exploring the advantages of using avatars as company representatives on e-commerce websites. Andersson and Sobek (2020) conducted a research study on virtual avatars and virtual influencers, examining the authenticity of avatars and proposing a model incorporating factors like goal, personality, continuity, and transparency. Lancere de Kam and Diefenbach (2020) focused on the online purchase intentions of Generation Y and the use of customized avatars in the fashion domain. Bleier et al. (2019) address the importance of creating effective online experiences for customers in their study. They highlight that the success of online retailing relies heavily on the creation of compelling online experiences through well-designed websites. However, the specific appearance of such web pages remains uncertain. This research investigates, through 16 experiments, how 13 unique design elements shape the four dimensions of online customer experience (information provision, entertainment, social presence, and sensory appeal), thus influencing purchase behavior. The study examines how product features and brand names amplify or diminish uncertainty associated with online purchases. A field experiment manipulating real product pages on the Amazon website confirms these findings. The results of this study provide guidance for managers in designing websites to enhance customer experiences. Sajadi's (2016) study delves into the effects of online interaction on users' inclination to explore and purchase products or services using avatars. The author acknowledges the growing prevalence of online interaction facilitated by avatars in contemporary settings. The study systematically investigates the historical context of online interaction, with a specific focus on trust, emotional appeal, and the presence of avatars. Selecting companies that incorporate avatars on their websites, the researchers invited participants to visit these websites and complete a comprehensive questionnaire. A total of 945 questionnaires were collected, allowing for a robust analysis of the data. The findings of this study underscore the significant positive impact of online interaction on purchase intention. Additionally, the research highlights the noteworthy influence of avatar presence on the development of trust and emotional appeal. This study contributes to the existing literature by shedding light on the intricate dynamics between avatars, online interaction, and users' purchasing behavior. In their study titled "The Effects of Avatars on Trust and Purchase Intention of Female Online Consumers: Consumer Knowledge as a Moderator," Lee et al. (2015) addressed the focus of recent studies on online consumer trust in websites or online retailers. However, customer attitudes may vary based on the type of avatar used since consumer trust can be cognitive or affective. Additionally, customers may rely on their product knowledge for judgment and decision-making. An online survey was conducted on female consumers with two scenarios related to skincare products, and 599 responses were used for hypothesis testing. The results indicate that avatars can indeed enhance customer trust in e-commerce. It was also evident that different avatar types motivate customer trust. Customers have higher trust in an expert avatar compared to an attractive avatar, while customers have higher affective trust in an attractive avatar compared to an expert avatar. Furthermore, avatars are effective for customers with varying levels of product knowledge, with the influence of avatar type being stronger for customers with high product knowledge. The impact of affective trust on purchase intention is stronger than cognitive trust. For skincare products, attractiveness may be more important than expertise in avatar design. In their research, Karimi et al. (2015) examined how individual decision-making styles and product knowledge influence the online buying process. They presented a typology of online buying decision-making behavior and identified four patterns of online consumers. The intensity of the decision-making process was found to be influenced by decision-making style and product knowledge. McGloin et al. (2014) explored the influence of avatars on perceptions of trustworthiness in the online consumer environment. They found that both the perceived avatar and textual credibility positively impact the perception of source trust. The disparity between perceived avatars and textual credibility can significantly influence the perception of source trust. Mennecke and Peters (2013) discussed "mavatars," new avatars used for marketing and consumer profile visualization. They presented a framework for understanding the utilization of mavatars and found that mavatars have the potential to create new options for personalized advertisements and relevant content, resulting in a more positive impact on consumers compared to regular advertisements.

In their study, Wang and Fodness (2010) propose and examine a model that demonstrates the influence of avatars as perceived social agents on consumers' internal evaluations and behavioral intentions within an online retail website. This study investigates both the social effects of avatars and their similarity to human sales personnel on consumer behavior. Overall, consumers exhibit more positive emotional responses and higher trust in an online retail website featuring a highly likable avatar compared to conditions where the avatar is either text-based or less likable. Simulated websites were utilized in this research to test the S-O-R hypotheses. The results indicate that avatars can successfully emulate interpersonal interactions with online customers and support and extend the social effects exerted on them Jin and Bolebruch (2009) investigated avatar-based advertising in virtual worlds, focusing on the impact of avatar presence and physical characteristics on consumer interaction, attitudes towards products, and enjoyment of the online shopping experience. Holzwarth et al. (2006) conduct a study titled which addresses the impersonal nature of online retailing as a barrier. The solution proposed to mitigate this issue involves utilizing avatars to provide product information. Avatars are computer-generated graphical representations that can be created using advanced technology. The study demonstrates that avatars can enhance the effectiveness of web-based sales channels. Specifically, incorporating an avatar into the existing website information leads to increased customer satisfaction with the seller, a positive attitude towards the product, and purchase intention. Furthermore, the research reveals that efforts to enhance the attractiveness and expertise of avatars can be persuasive to certain buyers. Overall, avatar attractiveness proves effective across all levels of engagement, while specialization of avatars yields positive outcomes primarily at higher levels of engagement. Additionally, the study highlights that an appealing avatar serves as an influential factor in persuading consumers to make purchases.

Hemp (2006) argued for the marketing potential of avatars in online interactions, emphasizing the opportunity for consumers to experiment with alternative identities. Avatars offer a unique and highly developed individual representation, allowing for sustainable interaction with brands and potential reassessment of marketing strategies.

Several previous studies, including Holzwarth et al. (2006), Lee et al. (2015), Sajadi (2016), and Gammoh et al. (2018), have examined aspects of the current research model regarding the impact of avatar presence and the role of entertainment and information provision. However, none of these studies have specifically investigated the variables of avatar presence and their impact on the purchase experience, the influence of the purchase experience on the intention to use avatars, and ultimately the purchase intention in digital marketing. Furthermore, considering the significance of digital marketing in today's world and the lack of knowledge in the field of avatar marketing, this research aims to analyze the influence of avatar presence on customers' online experience and their inclination to use avatars and make product purchases, with the mediating role of entertainment and information provision. This analysis can provide valuable insights into consumer behavior regarding avatar presence and its impact on the inclination to engage in online purchases.

As a result, the following hypotheses were proposed:

H1: The presence of avatars on social networks has a significant impact on customers' online experience.

H2: Entertainment mediates the relationship between avatar presence and customers' online experience.

H3: Information provision mediates the relationship between avatar presence and customers' online experience.

H4: Customers' online experience has a significant impact on the intention to use avatars.

H5: The intention to use avatars has a significant impact on the purchase intention.

RESEARCH METHODOLOGY

Research Model

Based on the hypotheses presented in the previous section, the research model has been constructed as illustrated in Figure 1.

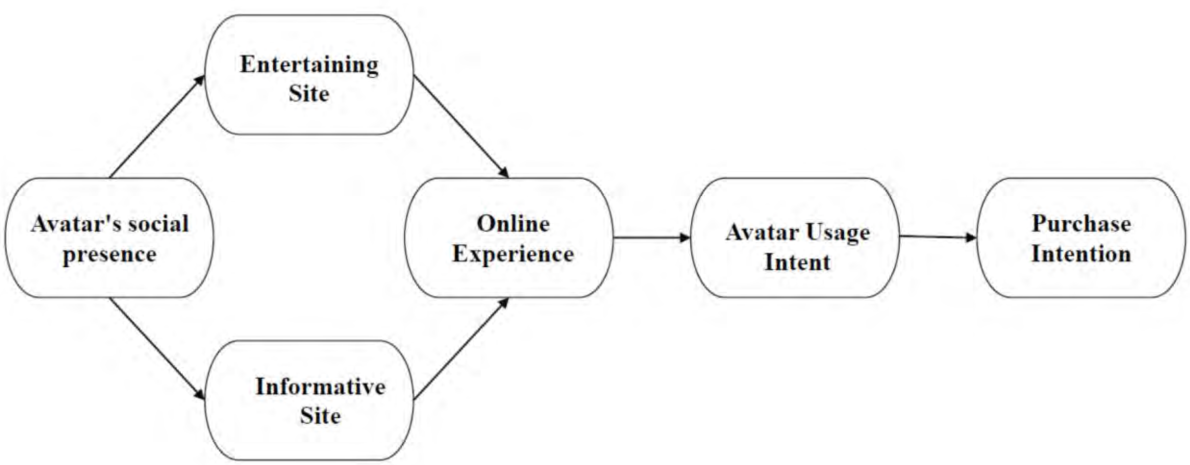


Figure 1: Research Model

Measures

This study examined six variables, each supported by different studies and assessed through a set of questions. The variable "Avatar Presence" was evaluated based on three studies: Holzwarth et al. (2006), Sajadi (2016), and Bleier et al. (2019), utilizing a total of three questions. The variable "Entertainment" was investigated using two studies: Holzwarth et al. (2006) and Sajadi (2016), involving a total of five questions. Similarly, the variable "Information Provision" was analyzed with three studies: Holzwarth et al. (2006), Sajadi (2016), and Wang and Fodness (2010), comprising a total of six questions. The variable "Online Experience" was examined based on three studies: Holzwarth et al. (2006), Sajadi (2016), and Bleier et al. (2019), with a total of three questions. The variable "Intention to Use Avatars" was investigated through three studies: Holzwarth et al. (2006), Sajadi (2016), and Lancere de Kam and Diefenbach (2020), using a total of six questions. Finally, the variable "Purchase Intention" was evaluated through two studies: Sajadi (2016) and Elwalda et al. (2016), involving a total of three questions.

Reliability and Validity of Measures

The questionnaire used in this study was examined and validated from the perspective of face validity by several experts in the field of marketing. The reliability of the questionnaire was determined using the Cronbach's alpha method, with a coefficient value of 0.863 for the entire questionnaire. Additionally, the Kolmogorov-Smirnov test, as well as the skewness and kurtosis tests, were employed to assess the normality of the questionnaire items, confirming the normal distribution of responses in the sample.

Sampling and Data Collection Procedures

The study population of this research consisted of individuals familiar with avatars, as well as those who had experience purchasing through avatars or followed popular avatars in virtual space. Due to the unavailability of a specific list of individuals familiar with avatars, a non-random convenience sampling method was employed. To collect the required data for this research, both library research and field research methods were used. Initially, the questionnaire items were extracted. Considering the use of the structural equation modeling approach by Schumacher and Lomax (2004) and the formula $5q<n<15q$ (q represents the number of questionnaire items, and n represents the sample size), a minimum of 130 samples is required for this research to accurately represent the study population. Over a period of 3 months, 390 questionnaires were shared and sent to the intended individuals. Out of this number, 370 questionnaires were validated and used for analysis. The questionnaire used in this study consists of three sections: introduction, general questions, and specific questions. The general questions include 5 items regarding the respondents' general information, such as age, gender, education, etc. The specific questions consist of 26 items related to avatar presence, the role of entertainment and information, online experience, intention to use avatars, and purchase intention.

DATA ANALYSIS AND RESULTS

Regarding the validity of the measurement model constructs, the obtained results consistently indicate that the factor loadings for all variables are greater than 0.5, demonstrating significant factor loadings. Consequently, the existing variables effectively measure the latent constructs. The final research model and the path coefficients between the model constructs are depicted in Figure 2. Based on the results of the overall fit indices, it can be concluded that the research model exhibits a good fit, and the proposed factorial structure is acceptable. The analysis results can be found in Table 1.

Table 1: General Fit Indices for Measurement Models				
Fit Index	Abbreviation	Value	Acceptable Fit	Status
Standardized Chi-Square	CMIN	494.2	1<….<3	Desirable
Root Mean Square Residual	RMR	0.240	> 0.050	Desirable
Goodness of Fit Index	GFI	0.917	> 0.900	Desirable
Adjusted Goodness of Fit Index	AGFI	0.906	> 0.900	Desirable
Normed Fit Index	NFI	0.936	> 0.900	Desirable
Relative Fit Index	RFI	0.927	> 0.900	Desirable
Incremental Fit Index	IFI	0.912	> 0.900	Desirable
Comparative Fit Index	CFI	0.909	> 0.900	Desirable
Root Mean Square Error of Approx.	RMSEA	0.075	< 1.000	Desirable

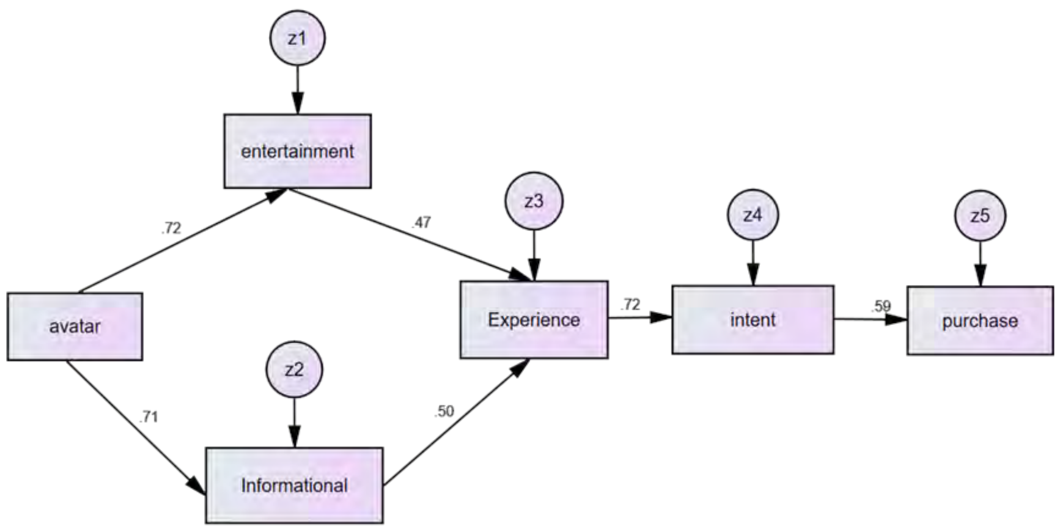


Figure 2: Research model with path coefficients.

Based on the hypothesis testing conducted and the results presented in Table 2, it can be confidently stated, with a 95% confidence level, that the presence of avatars on websites and social networks has an impact on customers' online experience. Furthermore, communication and entertainment act as mediating variables that influence the relationship between avatar presence and customers' online experience. Additionally, customers' online experience significantly influences their intention to use avatars. Moreover, the intention to use avatars has an impact on the purchase intention.

Table 2: Research Hypotheses Test Results				
Hypothesis	Estimate	S.E.	C.R.	P
Avatar's social presence → Informative site	.708	.039	12.516	***
Avatar's social presence → Entertaining site	.720	.054	13.380	***
Informative site → Online Experience	.500	.115	9.193	***
Entertaining site → Online Experience	.470	.061	8.438	***
Online Experience → Avatar Usage intent	.722	.059	13.864	***
Avatar Usage intent → Purchase intention	.586	.077	9.813	***

*** values less than 0.001

DISCUSSION AND CONCLUSIONS

The findings of this study, using the novel model presented in previous sections, align with previous research, providing further support for the hypotheses under investigation. Avatars on websites and social networks have a significant impact on customers' online experience. This finding aligns with the research by Sajadi (2016) and Holzwarth et al. (2006), which emphasize the positive effects of avatars on communication, trust, and social interaction in online shopping. The role of entertainment in enhancing the online experience is confirmed by Bleier et al. (2019), Sajadi (2016), and Wang and Fodness (2010). Avatars' social presence also influences entertainment and emotional value, as discussed by Holzwarth et al. (2006). The impact of avatar information on customer purchasing decisions and trust is supported by Blair et al. (2019), Sajadi (2016), and Holzwarth et al. (2006). Information plays a stronger role in online interaction compared to entertainment. Positive online shopping experiences lead to an increased intention to use avatars, as confirmed by Lancere de Kam and Diefenbach (2020), Sajadi (2016), and Holzwarth et al. (2006). The influence of avatars on the intention to purchase is supported by Sajadi (2016), Holzwarth et al. (2006), McGloin et al. (2014), Li et al. (2015), and Elwalda et al. (2016). Avatars enhance customer satisfaction, product attitudes, and trust, thereby increasing purchase intentions.

In today's digital landscape, avatars have emerged as a powerful tool for enhancing customers' online experience across various industries. To optimize this impact, online businesses should strategically incorporate avatars on their websites, tailoring their appearance to match the specific field and product being offered. For specialized product sellers, utilizing an avatar that resembles an expert in the field can provide consumers with valuable information and deeper engagement. Avatars should engage in intimate conversations with customers, acting as sales representatives, establishing connections, and ensuring enjoyable interactions. Employing artificial intelligence to power avatars enables seamless communication without the need for human guidance, allowing them to provide answers, guidance, and entertainment. Information dissemination is a critical factor influenced by avatars, particularly for service, banking, and educational sectors, which should use unique avatars to present services transparently, step-by-step, and even incorporate online voice conversations. The use of exclusive avatars on university websites can facilitate new student onboarding and aid international students with language support. Custom avatars play a vital role in well-known businesses, as they can be personalized based on customer behavior and purchase history, delivering tailored product recommendations and services that lead to a memorable online experience, fostering long-term customer loyalty. Furthermore, businesses aiming to establish a positive image and brand identity can leverage avatars as virtual influencers, creating personal pages where they share their experiences with products and services, thereby solidifying their position as trusted advisors and influencers. Overall, by leveraging the potential of avatars strategically, businesses can elevate customers' online experiences, foster engagement, and drive meaningful connections, ultimately leading to increased intention to purchase and long-term success.

The integration of avatars in online customer experiences offers several notable advantages. Firstly, avatars bridge the gap between the impersonal nature of online interactions and the desire for social connection, enhancing customers' engagement and emotional involvement with websites and brands. Additionally, avatars can serve as personalized virtual sales representatives, offering tailored product recommendations and assistance that cater to individual customer needs. This level of customization enhances customer satisfaction and encourages repeat purchases, fostering brand loyalty. Moreover, avatars contribute to an interactive and dynamic user experience, making the online shopping process more enjoyable and immersive. They can provide real-time responses to inquiries, reducing the need for customers to wait for human support, thus expediting decision-making and purchase actions. However, there are potential drawbacks to consider when implementing avatars in online experiences. One challenge

is ensuring that avatars maintain a consistent and appropriate tone, appearance, and messaging across various interactions to avoid misunderstandings and discrepancies. Over-reliance on avatars might lead to a lack of personal touch, which could deter some customers who prefer human interactions. Additionally, the effectiveness of avatars can vary depending on cultural differences and individual preferences, potentially resulting in misinterpretation or disengagement. Moreover, the integration of avatars necessitates technological infrastructure and maintenance, which could be costly and time-consuming for businesses. Lastly, there's a concern about the potential for avatars to be perceived as gimmicky or insincere, especially if not implemented thoughtfully.
This study discussed and investigated the impact of avatars on customers' online experience and their inclination to use avatars and purchase products, considering the mediating role of entertainment and information. The results of hypothesis testing showed that the presence of avatars on websites and virtual networks had a positive and significant impact on customers' online experience, with information and entertainment acting as mediators. Furthermore, it was found that customers' intention to use avatars also influenced their purchase intention. Based on the limitations identified in this study, several suggestions are proposed for future researchers in this field. It is suggested that they investigate additional avatar features such as appearance, gender, unique facial characteristics, and sense of humor, examining their importance and impact on customer trust. Additionally, it is recommended to measure the influence of avatars on customer loyalty, given its significance to brand loyalty. Furthermore, it is advised to conduct separate research to examine the impact of avatars on customers' digital experience. It is also suggested that, considering recent advancements in the field of artificial intelligence, the possibility of individually customizing avatar identities for customers based on their previous purchasing information be investigated.

REFERENCES

Andersson, V., & Sobek, T. (2020). Virtual Avatars, Virtual Influencers & Authenticity.

Bleier, A., Harmeling, C. M., & Palmatier, R. W. (2019). Creating effective online customer experiences. Journal of marketing, 83(2), 98-119.

Elwalda, A., Lü, K., & Ali, M. (2016). Perceived derived attributes of online customer reviews. Computers in Human Behavior, 56, 306-319.

Etemad-Sajadi, R., & Ghachem, L. (2015). The impact of hedonic and utilitarian value of online avatars on e-service quality. Computers in Human Behavior, 52, 81-86.

Etemad-Sajadi, R. (2016). The impact of online real-time interactivity on patronage intention: The use of avatars. Computers in human behavior, 61, 227-232.

Gammoh, B. S., Jiménez, F. R., & Wergin, R. (2018). Consumer attitudes toward human-like avatars in advertisements: The effect of category knowledge and imagery. International Journal of Electronic Commerce, 22(3), 325-348.

Hemp, P. (2006). Avatar-based marketing. Harvard business review, 84(6), 48-57.

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. Journal of marketing, 70(4), 19-36.

Jin, S. A. A., & Bolebruch, J. (2009). Avatar-based advertising in Second Life: The role of presence and attractiveness of virtual spokespersons. Journal of Interactive Advertising, 10(1), 51-60.

Karimi, S., Papamichail, K. N., & Holland, C. P. (2015). The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour. Decision Support Systems, 77, 137-147.

Kravchik, I., & Newhall, E. (2020). Avatars Engage Generation Z. Journal of Strategic Innovation and Sustainability, 15(5), 55-61.

Kumar, P. (2020). Influence of Avatar Features and Its Presence on Customer's Behaviour: Literature Review. Journal of Research in Business and Management, 8(11), 23-28.

Lancere de Kam, E., & Diefenbach, J. (2020). Understanding the Digital Future: Applying the Decomposed Theory of Planned Behaviour to the Generation Y's Online Fashion Purchase Intention while Creating and Using a Customised Avatar.

Lee, H. S., Sun, P. C., Chen, T. S., & Jhu, Y. J. (2015). The effects of avatar on trust and purchase intention of female online consumer: Consumer knowledge as a moderator. International Journal of Electronic Commerce Studies, 6(1), 99-118.

McGloin, R., Nowak, K. L., & Watt, J. (2014). Avatars and Expectations: Influencing Perceptions of Trustworthiness in an Online Consumer Setting. PsychNology Journal, 12.

Mennecke, B. E., & Peters, A. (2013). From avatars to mavatars: The role of marketing avatars and embodied representations in consumer profiling. Business Horizons, 56(3), 387-397.

Molina, L. (2021). Celebrity Avatars: A Technical Approach to Creating Digital Avatars for Social Marketing Strategies (Doctoral dissertation, Florida Atlantic University).

Moustakas, E., Lamba, N., Mahmoud, D., & Ranganathan, C. (2020, June). Blurring lines between fiction and reality: Perspectives of experts on marketing effectiveness of virtual influencers. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security) (pp. 1-6). IEEE.

Park, S., Choi, W. J., & Shin, D. (2021). Who Makes You More Disappointed? The Effect of Avatar Presentation, Company Market Status, and Agent Identity on Customers' Perceived Service Quality and Satisfaction of Online Chatting Services. Asia Marketing Journal, 23(3), 4.

Silva, M. J., de Oliveira Ramos Delfino, L., Alves Cerqueira, K., & de Oliveira Campos, P. (2022). Avatar marketing: A study on the engagement and authenticity of virtual influencers on Instagram. Social Network Analysis and Mining, 12(1), 130.

Wang, L. C., Baker, J., Wagner, J. A., & Wakefield, K. (2007). Can a retail web site be social? Journal of marketing, 71(3), 143-157.

Wang, L. C., & Fodness, D. (2010). Can avatars enhance consumer trust and emotion in online retail sales? International Journal of Electronic Marketing and Retailing, 3(4), 341-362.

DECISION SCIENCES INSTITUTE

Analyzing the interplay between big data analytics-enabled dynamic capabilities and contextual factors in innovation.

Adilson Carlos Yoshikuni
Technology & Analytics in Management, Accounting, and Finance (TAMAF)
Mackenzie Presbyterian University
Email: adilson.yoshikuni@mackenzie.br

Marcio Santos Quadros
Technology & Analytics in Management, Accounting, and Finance (TAMAF)
Mackenzie Presbyterian University
Email: marcio.quadros@gmail.com

Rajeev Dwivedi
Eastern Washington University
Technology & Analytics in Management, Accounting, and Finance (TAMAF)
Mackenzie Presbyterian University

Email: rajeevdwivedi@gmail.com

Pradeep Dwivedi
Technology & Analytics in Management, Accounting, and Finance (TAMAF)
Mackenzie Presbyterian University
Email: softwarepra@gmail.com.

ABSTRACT

The researchers and practitioners are urged to understand the role of big data analytics (BDA) in enabling dynamic capabilities (DC) to create innovation under organizational factors and which combinations of conditions. To address the issues, it was examined the fit between the BDA-DC interplay with organizational contextual factors in driving a firm's innovation. Survey data from 188 US firms are analyzed using fuzzy set qualitative comparative analysis (fsQCA). Results demonstrated that under different combinations of contextual factors, the significance of BDA-DC varies, with specific configurations leading to high levels degree of innovation.

KEYWORDS: Exploration, exploitation, innovation, big data analytics-enabled dynamic capabilities, organizational factors, fuzzy set qualitative comparative analysis.

INTRODUCTION

Organizations should increase product and service innovations to be competitive under the current economic environment characterized by higher uncertainty and turbulence, hyper-competition, and global and local markets (Mikalef *et al.*, 2020a; Wamba *et al.*, 2020; Dwivedi *et al.*, 2023; Yoshikuni *et al.*, 2023).

Thus, environmental uncertainty, turbulence, and competition increased, and firms need to develop organizational capabilities to quickly identify opportunities and react to threats and market challenges through innovation (Ravichandran, 2018; Ilmudeen, 2022). Uncertainty can be characterized by external factors that affect firms through the amount of information required to perform a task to gain a competitive advantage. The factors of an uncertain environment are commonly described in three dimensions: dynamism, heterogeneity, and hostility (Newkirk and Lederer, 2006; Yayla and Hu, 2012; Yoshikuni and Lucas, 2022)

Teece and colleagues (Teece, Peteraf and Leih, 2016; Teece, 2018)) describe that an organization must survive and lead with external factors to build dynamic capabilities in innovation into sensing, seizing, and transforming routines to attend market requirements.

Davenport and colleagues (Davenport, Harris and Morison, 2010) argue that big data analytics can provide the ability for firms to collect data and generate actionable insight from large volumes of unstructured data to lead to higher uncertainty and turbulence environment. Braganza and colleagues (Braganza *et al.*, 2017) discussed to firms to create significant innovation in a volatile and dynamic environment, and they need to develop dynamic capabilities by emerging technologies as big data analytics.

Current studies have demonstrated the role of big data analytics in enabling dynamic capabilities in sensing, seizing, and transforming capabilities to gain innovation in the competitive market under uncertainty environment (Conboy *et al.*, 2020; Mikalef *et al.*, 2020a; Wamba *et al.*, 2020).

The study aims to identify in what ways BDA capabilities with organizational factors and the environment towards the enhancement of innovation. As organizations compete in different contexts through specific firm's characteristics, and the BDA capabilities developed will be shaped to fit the environment in which they operate.

Based on the goals of BDA are targeted towards developing management's ability to create different combinations to drive firm innovation (Van de Wetering, Mikalef and Krogstie, 2019; Mikalef and Krogstie, 2020). This research defines dynamic capabilities as a firm ability to reallocate and reconfigure its resources and capabilities by significantly improving them to capitalize on market opportunities and develop operational adjustments to address customers' needs (C.E. Helfat and Raubitschek, 2018).

Furthermore, the study examines organizational and environmental factors. The organizational factors involve aspects of competitive strategy [low cost and differentiation (Porter, 1998)] size and sector of firms (Melville, Kraemer and Gurbaxani, 2004). The external environment captures the dynamism, hostility, and heterogeneity of external factors in which organizations operate and compete. Therefore, the follow research question is proposed in this study:

RQ. What are the combinations among big data analytics-enabled dynamic capabilities and context factors are relevant for creating innovation?

Thus, through the lens of the firm's Resource-Based View of theory (RBV), it addresses the research question to identify the relevant BDA capabilities and contingency theory to investigate how contextual factors coalesce with the dynamic capabilities to drive innovation. Building on 188 responses from USA firms through the fsQCA methodological tools. The fsQCA examines the patterns of elements that lead to high levels of innovation. Hence, the tools allow identifying the non-linear relationships between BDA capabilities, contextual factors, and their effect on explorative and exploitative innovation.

The study is structured through the central concepts of the relevant BDA capabilities, which are grounded on the RBV and dynamic capabilities view (DCV). The conception of innovation and the role of contextual factors is discussed through the management literature and contingency theory lens.

The following section describes the methodology employed, demonstrating the gathering data, measurements, reliability, and validity tests, the fsQCA analyses and results by sensitivity and predictive validity tests. The final section shows the study's theoretical and practical implications.

THEORY DEVELOPMENT

Resource-Based View and Dynamic Capabilities View

The RBV studies of the firm has been investigating in the last four decades (Helfat, Kaul and Ketchen, 2023). The RBV literature refers to the role of organizational capabilities to gain competitive advantage through the degree to which they are valuable, rare, cannot be perfectly reproduced, and are non-substitutable in unique combinations (Amit and Schoemaker, 1993; Barney, 2001).

Recent studies in the strategic management literature have shown the difference between operational or ordinary capabilities and dynamic capabilities (Eisenhardt and Martin, 2000; Teece, 2007; Constance E. Helfat and Raubitschek, 2018). Ordinary capabilities allow organizations to perform an activity using more or less the same resources to support existing products and services to meet the same customer need (Helfat and Winter, 2011). Dynamic capabilities enable firms to build the ability to reconfigure and change in response to the challenge of the external environment (Winter, 2003; C.E. Helfat and Raubitschek, 2018).

Previous studies have demonstrated that causal mechanisms through which dynamic capabilities increase their value, and it is amplified under conditions of higher environmental challenge (Mikalef and Pateli, 2017; Wamba *et al.*, 2020; Yoshikuni, 2022). In this study dynamic capabilities refer to the firm capacity to purposefully reconfigure and deploy its organizational capabilities to attend to changing business requirements (Teece, 2007).

Moreover, past research on dynamic capabilities have empirically demonstrated that organizational capabilities affect proximate and distal performance outcomes in multiple ways; by facilitating traditional and new business models (A. Jantunen *et al.*, 2018), contributing to marketing and operational capabilities (Lu and Ramamurthy, 2011; Mikalef and Pateli, 2017; Steininger *et al.*, 2022), and enabling innovation and improvisational capabilities (Yoshikuni and Dwivedi, 2023).

These studies on dynamic capabilities confirmed that firms' operational capabilities could affect proximate and distal innovation outcomes. Thus, in line with Teece and colleagues (Teece, Peteraf and Leih, 2016), this study refers to the dynamic capabilities in innovation "as the firm’s capacity to innovate, adapt to change, and create change that is favorable to customers and unfavorable to competitors."

Hence, based on the conceptual definition of Teece and colleagues (Teece, Peteraf and Leih, 2016) and past empirical studies (Mikalef *et al.*, 2020a; Wamba *et al.*, 2020), this study adopted three primary organizational capabilities: (a) sensing capabilities to identify, develop, co-develop, and assess opportunities (and threats) related to market and customer requirements, (b) seizing capabilities to capture and address opportunities by mobilization of firms resources, (c) renewing and transforming business processes to create innovation and attend customers need.

Big Data Analytics

As stated, the role of organizational capability, IT capability refers to the capacity to firm mobilize and deploy IT-based resources to enable business strategies and business processes (Bharadwaj, 2000). IT capabilities in previous studies are defined as three main types of resources, in line with the organizational resource of strategic management literature (Barney, 1991). These include tangible IT resources, human IT resources, and intangible IT-enabled resources (Bharadwaj, 2000; Mikalef and Pateli, 2017; Li and Chan, 2019; Yoshikuni and Dwivedi, 2023).

This study emphasizes that IT is embedded in organizational capabilities, enabling dynamic capabilities to leverage proximate and distal outcomes under environmental changes Thus, the research aims to understand how IT manifests business value through emerging technologies. Current research has been developing a consensus that big data analytics (BDA) capabilities enable organizational processes through dynamic capabilities (DC) in innovation (Mikalef *et al.*, 2020a; Wamba *et al.*, 2020).

In this study, the potential IT business value examines the BDA capabilities in combination or co-present with other complementary organizational resources that firms must build to be able to leverage DC in innovation. Hence, based on the notion of strategic BDA business value, big data analytics-enabled dynamic capabilities (BDA-DC) are defined as a "firm's abilities to sense, seize and transform capabilities enabled through big data and business analytics, in orchestration with other organizational resources and capabilities, to leverage innovation and respond to business environmental challenges" (Yoshikuni and Dwivedi, 2022, p. 44).

BDA-DC in terms of the sensing ability consist in detect market requirements and unrecognized customer need, support managers to understand their clients' principal requirements (Mikalef *et al.*, 2020b). Wamba and colleagues (Wamba *et al.*, 2020) demonstrated how BDA capabilities support firms to sense rapid changes in the factors' environment to enable managers to make immediate plans to respond to the challenges of an uncertain environment.

Thus, BDA-DC has the processing power to capture large volumes at high velocity and a variety of data to transform raw data into actionable insight, improving response speed in shorter cycle times to deal with environmental changes (Dong and Yang, 2018; Popovič *et al.*, 2018).

Seizing capabilities enabled by BDA allows firms to deploy strategic analytics knowledge to examine and evaluate customer behaviors with the aim to incremental and develop new products tailored to customers' expectations (Ciampi *et al.*, 2021). Therefore, BDA-DC in terms of seizing capabilities ability firms to draft, evaluate and select various potential solutions to ultimately create a plan to carry out a potential solution (Mikalef *et al.*, 2020b).

Hence, BDA can process the large and abundant data from market requirements, prioritizing, categorizing, selecting, and ranking customer preferences, improving the manager's quality and efficiency in decision-making to incremental and create new products and services (Dong and Yang, 2018) under higher challenges of external factors (Wamba *et al.*, 2020).

According to Mikalef and Krogstie (2020) and Ciampi and colleagues (Ciampi *et al.*, 2021), BDA has an essential role in building organizational capabilities to firms create disruption and incremental innovation in a higher turbulence environment.

Innovation

Previous studies have argued that innovation is a vital performance outcome for a firm to compete and survive in higher environmental challenges The innovation is the performance outcome that demonstrate the successful implementation of creative ideas by a firm (Benner and Tushman, 2015; Blank and Naveh, 2019).

To attend to market requirements and customer segments, firms need to leverage different degrees of innovation, ranging from exploitative to exploratory innovation (Jansen *et al.*, 2006; Yoshikuni and Lucas, 2022).

Exploitative innovation refers to create innovation by expanding existing products and services to meet needs of customers and markets requirements (Jansen *et al.*, 2006). Hence, exploitative innovations are considered the incremental innovations of products and services, which are developed to meet the needs of existing customers or market (Zhang *et al.*, 2023). Organizations use existing abilities to improve and increase the efficiency of routines and business processes by existing knowledge and skills (Xie *et al.*, 2020; Yoshikuni and Lucas, 2022). Therefore, exploitative innovations build on existing abilities of skills and competence and existing resources, structures, and processes to create incremental innovation (Benner and Tushman, 2015; Albors-Garrigos, Igartua and Peiro, 2018).

Exploratory innovation refers to providing new products and services to meet the needs of emerging customers or markets (Jansen *et al.*, 2006). Thus, firms build abilities to create products and services offering new designs to satisfy the needs of potential customers and markets through creating and developing new markets and channels (Benner and Tushman, 2015; Blank and Naveh, 2019). Hence, organizations need new knowledge, skills, competence, and abilities to promote ideas that impact in radical changes to their products and services to meet the needs of potential customers (Benner and Tushman, 2015; Albors-Garrigos, Igartua and Peiro, 2018).

.

Contextual factors

The contingency theory assumes operational capabilities can be varied through different aspects of management practices and resources influenced by the external environment (Donaldson, 2014). The performance outcomes are simultaneously influenced by numerous contingency forces, and these kinds of factors might complement or counteract one another (Wolf and Floyd, 2017).

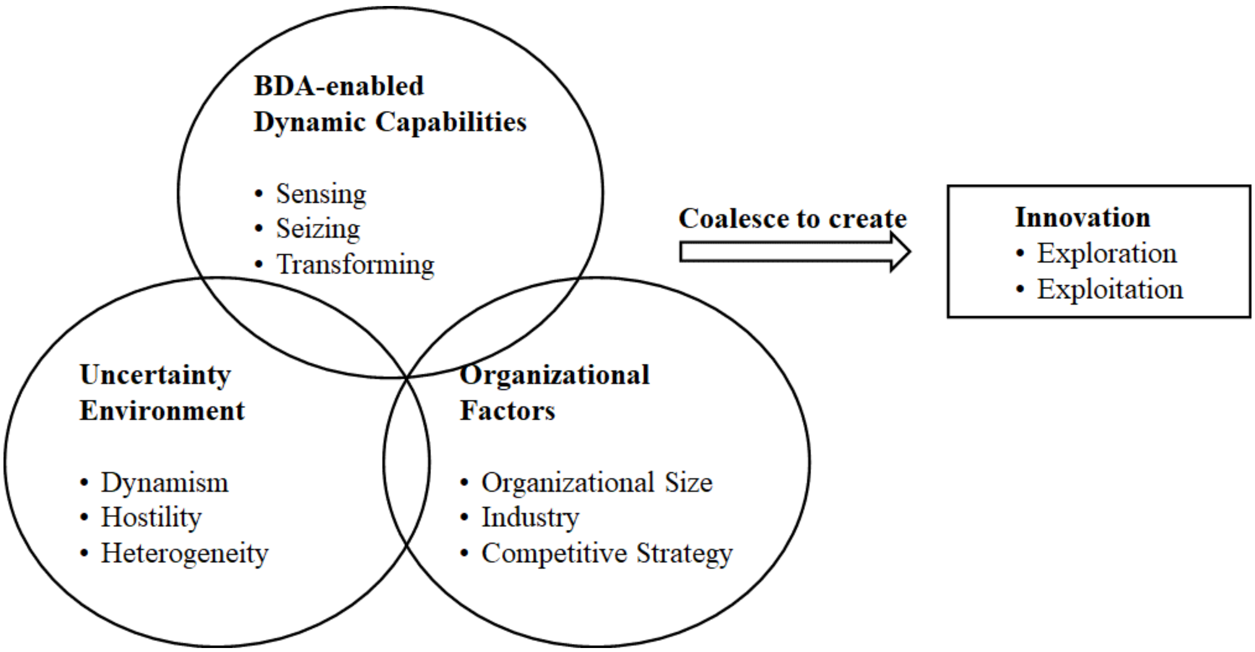
Moreover, the competitive strategy (low cost and differentiation), value chain of a firm and the adoption of technologies interplay among different contextual (or contingency) factors (Porter, 1998). Melville and colleagues (Melville, Kraemer and Gurbaxani, 2004) argue that external and internal factors can influence the adoption and operational capabilities through resources of tangible IT, human IT, and intangible IT-enabled organizational resources. According the researchers (Melville, Kraemer and Gurbaxani, 2004), there is evidence of structural differences across industries regarding the ability to use IS for improve proximate and distal outcome performance under competitive environment.

Previous studies have examined the influence of industry characteristics in the competitive environment, linked BDA capabilities enabling dynamic capabilities to create innovation (Mikalef and Krogstie, 2020; Wamba *et al.*, 2020; Ciampi *et al.*, 2021).

Research framework

As stated, firm contextual factors and external factors of the environment can condition the effect of dynamic capabilities of BDA and their subsequent importance towards creating innovation gains. Figure 1 depicts the interplay among BDA-DC, uncertainty environment, and organizational factors in creating innovation. The figure shows the interactions between all the elements and suggests that specific causal recipes will create innovation.

Figure 1.
Research framework of configurations for creating innovation



RESEARCH METHODS

Scale

The research model was designed using current literature constructs, and novel constructs were developed and validated by a new scale.

Based on current literature and tests with executives and scholars was created the new multi-item scale of BDA-enabled dynamic capabilities (BDA-DC) construct. BDA-DC as Type II second-order construct [reflective first-order, formative second-order (Hair *et al.*, 2017)] and it composed through three reflective first-order dynamic capabilities constructs (Teece, Peteraf and Leih, 2016).

The tests of convergent and discriminant validity of the first-order BDA-DC was adapted measures of (a) sensing, (b) seizing, and (c) transforming (Mikalef and Krogstie, 2018; Wamba *et al.*, 2020; Yoshikuni and Dwivedi, 2022) and strategic management literature (Porter, 1998; Mintzberg, Ahlstrand and Lampel, 2009; Teece, Peteraf and Leih, 2016; Ari Jantunen *et al.*, 2018; Teece, 2018).

The BDA-DC construct was validated according to the guidelines of IS field (MacKenzie, Podsakoff and Podsakoff, 2011), in three steps: First, the measures' items of BDA-DC construct were generated through the authors. Second, the experts have assessed instrument content validity by qualitative and quantitative validation (Lewis, Templeton and Byrd, 2005). Third, six scholars and five practitioners (senior executives) realized the qualitative assessment by integrating of q-sort methodology and the Content Validity Ratio (CVR) techniques. Hence, as expected, certain questionnaire items were not grouped, so they were modified or removed of BDA-DC construct.

The constructs were adopted by current studies of competitive strategy [Low cost and differentiation (Porter, 1998)], uncertainty environment [Dynamism, hostility and heterogeneity (Newkirk and Lederer, 2006)], and characteristics of firm [Size and sector (Melville, Kraemer and Gurbaxani, 2004)]. All constructs were operationalized by Likert scale scores were coded from 1 to 7 (from 1 = “strongly disagree” to 7= “strongly agree”), see Appendix II.

Sample

The data was collected in US by convenience sampling data collection. The authors contacted the respondents in each firm using different sources and networks, such as personal contacts, professional association contacts, forums, mailing lists, and directories, in line with current research on empirical IS studies. The respondents were executives that know about management activities that involved activities related to all constructs. Moreover, the survey instructions asked respondents if they were not highly knowledgeable of specific information to consult other members.

Table 1 shows the demographic data and reveals that 53% (101) of the participants were senior executive managers. The industries were composed of manufacturing (42%), and services (58%). Competitive strategy concentrated in low cost with 117 (61%). The size firm was primarily represented by large size (59%).

Statistical technique

The research used the structural equation modeling based on the partial least square method (PLS-SEM) by SmartPLS version 3.3.3 because i) allow flexibility related to the assumptions on multivariate normality; ii) handling of structural model complexity and using smaller samples; iv) uses as a predictive statistical power tool for theory building (Hair et al., 2017).

Table 1
Demographic data

		Brazil	
Characteristics		Number	%
Respondent's position	Senior/executive manager	101	53%
	Middle/first line manager	87	47%
Competitive Strategy	Low Cost	117	61%
	Differentiation	74	39%
Firm size (number of employees)	Small size (1 to 499)	78	41%
	Large size (above 500)	113	59%
Industry sectors	Manufacturing	92	42%
	Services	99	58%

Measurement model

The test was conducted in reliability, convergent validity, and discriminant validity. The reliability at the level of the construct was assessed by examining Composite Reliability (CR) values were above the threshold of 0.70, indicating acceptable construct reliability (Fornell and Larcker, 1981), see Table 2. Convergent validity was examined whether AVE values were above the lower limit of 0.50, see Table 2.

The discriminant validity was examined in three ways. First, it was verified if each construct's AVE square root values are greater than its highest correlation with any other construct (Fornell-Larcker criterion). Second, it was tested if each indicator’s outer loading was greater than its cross-loadings with other constructs (Bido and Silva, 2019; Almeida *et al.*, 2022). The heterotrait-monotrait ratio (HTMT) was analyzed, and constructs their values were below 1 were remained. Heteregeneity construct was excluded. Thus, discriminant validity was satisfactory as recommended by Hair, Henseler and colleagues (Henseler, Ringle and Sarstedt, 2015; Hair *et al.*, 2017). Hence, all items were appropriate indicators for the respective latent variables, see Table 2 and Appendix I. The model fit was performed on composite-based standardized root mean square residual (SRMR), and the SRMR value is 0.069, which is below the threshold of 0.08, confirming the overall fit of the PLS path model.

Table 2
Assessment of convergent and discriminant validity

Constructs	1	2	3	4	5	6	7
1-Sensing	0,803						
2-Seizing	0,756	0,788					
3-Transforming	0,752	0,761	0,779				
4-Dynamism	0,621	0,613	0,678	0,759			
5-Hostlity	0,605	0,536	0,611	0,574	0,746		
6-Exploration	0,712	0,650	0,720	0,692	0,546	0,763	
7-Exploitaiton	0,655	0,542	0,583	0,606	0,619	0,726	0,780
Means	4,870	4,960	4,870	4,830	4,920	4,820	4,840
CR	0,879	0,867	0,860	0,803	0,832	0,847	0,861
AVE	0,645	0,621	0,606	0,576	0,556	0,582	0,608

Set-theoretic Analysis with fsQCA

This study follows a configurational perspective and uses a set-theoretic method called fuzzy-set qualitative comparative analysis (fsQCA) to explore how key elements come together to explain a particular outcome (Ragin, 2008). The research aims to determine which BDA-DC, organizational factors, and uncertainty environment are most significant in forming innovation of exploration or exploitation. The method of fsQCA adheres to the principles of configurational theories, which allow for the examination of interplays among messy and non-linear elements (Fiss, 2011; Pappas and Woodside, 2021).

Using the fsQCA method, it can identify the key factors and conditions contributing to developing strong explorative and exploitative innovation in firms. This approach differs from traditional variance-based methods that look for correlations between independent and dependent variables (Mendel and Korjani, 2012). Instead, fsQCA seeks out patterns of elements that lead to a specific outcome, reducing unnecessary factors and focus only on the necessary and sufficient conditions. Through this process, it can identify core and peripheral elements that contribute to the outcome, with core elements having a stronger causal relationship than peripheral elements (Fiss, 2011). This approach provides a more nuanced understanding of the factors that drive exploration and exploitation of innovation.

Through the research process of analyzing intricate connections among various components, and complex relationships between multiple elements, fsQCA presents several strengths (Ragin, 2008). This approach utilizes set theory, Boolean algebra, and counterfactual analysis to identify causal effects that lead to a desired outcome. This approach makes it possible to go beyond the limitations of conventional interaction term effects in regression analysis, which tend to be constrained to three-way interaction effects with a deeper understanding of the relationships between various elements and how they contribute to a specific outcome(Ragin, 2008; Pappas and Woodside, 2021).

Therefore, the fsQCA is a tool that supports mechanisms that help to deal with complex multidirectional interaction relationships, where all relevant elements that impact the result participate (Woodside, 2014). This tool can reduce the concern of an unobserved probability of heterogeneity and overcome some of the limitations of cluster analysis methods, which tend to seek clusters of homogeneous cases based on empirical quantitative data that lack theoretical foundations and control over the result. It is important to note that while cluster analysis methods can help identify groups of similar cases, they have limitations when explaining how those clusters are formed (Tóth *et al.*, 2015). That is where fsQCA comes in. Unlike traditional cluster analysis, fsQCA allows researchers to select the outcome they are interested in and relevant causal variables, producing multiple bundles that lead to that outcome. This approach enables a more in-depth examination of the specific combination of different elements that lead to desired outcomes. FsQCA is a valuable tool for researchers looking to understand the complex relationships between variables in achieving outcomes (Ragin, 2008; Pappas and Woodside, 2021)

Calibration

To use fsQCA is necessary to define both the outcome and independent measures. The next step is to calibrate all the measures into fuzzy sets with values from 0 to 1. The value 1 means full set membership, while a value of 0 implies no set membership. Therefore, all variables are continuous, with a scale ranging from 0 to 1, indicating the degree of membership. The direct method proposed by Woodside and colleagues (Woodside, 2014; Pappas and Woodside, 2021) was followed to transform continuous variables into fuzzy sets. For this, the fsQCA 4.0 software package was used, which converts a variable into a fuzzy set using the log-odds metric and the distance of the variable's value from the crossover point with the values of full membership and full non-membership as the upper and lower bounds. The degree of set membership is based on three anchor values: a full set membership threshold value (fuzzy score = 0.95), a full non-membership value (fuzzy score = 0.05), and the crossover point (fuzzy score = 0.50) (Ragin, 2008; Pappas and Woodside, 2021).

The current study utilizes a 7-point Likert scale to measure constructs, and in order to calibrate them into fuzzy sets, we have followed the guidelines put forth by (Fiss, 2011; Pappas and Woodside, 2021). Full membership thresholds have been set for values over 6.5, the crossover point has been set at 5, and full non-membership values at 3. These thresholds are consistent with other empirical studies that transform 7-point Likert scale variables into fuzzy sets (Woodside, 2014; Mikalef and Krogstie, 2020). These values were defined for the calibration of all variables measured in this study; several sensitivities analyses were performed to further validate our calibration.

The firms were classified based on their organization size, product or service-oriented industry, and organizational aspects of competitive strategy (low-cost strategy or differentiation strategy). These variables are assigned code 0 or 1 for large enterprises, small firms, product industry companies, and service industry ones, as the same process is applied to low-cost strategy and differentiation strategy. This was done in accordance with past empirical literature that uses mutually exclusive categories in the form of binary variables to simplify the analysis and reduce the number of possible remainder rows to make the results more understandable and represent the binary variables (Mikalef and Krogstie, 2020; Pappas and Woodside, 2021).

Analysis

This research investigated how a combination of BDA-DC, organizational factors, and an uncertainty environment can contribute to exploration and exploitation of innovation. This study produced results that offer various solutions for enhancing innovation. To achieve this, we conducted two separate fsQCA analyses, one for each dependent variable, (exploration and exploitation of innovation).

The analysis produced a 2k row truth table, where each row represented a possible combination or solution. Although previous studies recommend disregarding solutions with a consistency level lower than 0.80 (Ragin, 2008). This study considers a higher consistency level of 0.90, which means that the cases do not correspond to the expressed set theory relations into a solution (Fiss, 2011; Pappas & Woodside, 2021). In other words, the consistency score measures how reliably a combination produces results, similar to the significance level in standard econometric analysis (Woodside, 2014).

Following guidelines from prior studies, a minimum of three cases are defined for each solution (Woodside, 2014). Once these parameters are established, fsQCA analyses are then performed using exploration and exploitation as dependent variables. To further understand the results, truth tables for each outcome variable are presented, along with the number of cases of each composite solution.

The analysis method involves using Boolean algebra to simplify the truth table rows. We utilized the algorithm described by Ragin and colleagues (Ragin, 2008; Mendel and Korjani, 2012), which employs a counterfactual analysis of causal conditions to differentiate central and peripheral causes. Central causes are deemed crucial for the desired outcome, while peripheral causes can be viewed as less significant or replaceable (Fiss, 2011).

When using the fsQCA methodology, researchers consider core elements as strong causal relationships and peripheral elements as weaker predictors (Fiss, 2011; Pappas & Woodside, 2021). The algorithm generates three different results: parsimonious, intermediate, and complex solutions. However, to determine the central and peripheral elements, it is sufficient to examine the first two solutions, namely parsimonious and intermediate. The central conditions are those that are included in both of those solutions, whereas peripheral conditions are removed from the parsimonious solution and only appear in the intermediate solution according to Fiss (2011).

The outcomes of the fuzzy set analysis can be found in Table 3. The solutions are organized in vertical columns and marked with either black circles (●) to indicate the presence of a condition, crossed-out circles (ⓧ) to signify the absence of a condition, or white spaces to denote a "doesn't matter" situation. Large black circles are used to represent core conditions, while small circles indicate peripheral conditions. Each set of solutions is arranged according to the most important conditions.

Configurations for high innovation

Table 3 shows the solution of the fuzzy-set analysis results in five solutions for each type of innovation. Each solution exhibits consistency values above the set threshold and includes core and peripheral elements. Solutions analyzed consistency levels, raw and unique coverage, and overall solution consistency and coverage. Thus, consistency refers to the degree to which cases correspond to the set theoretic relationships expressed in a solution (Fiss, 2011).

The solutions provided in Table 3 for achieving high levels of explorative and exploitative innovation indicate the presence of core and peripheral conditions and neutral permutations for ten configurations. These ten solutions show high levels of consistency (consistency above 0.9) and explain a substantial number of cases (coverage above 0.4), indicating a substantial proportion of the outcome.

Table 3.

Configurations for high exploration and exploitation innovation.

	Solution									
	Exploration					Exploitation				
Configuration	1	2	3	4	5	1	2	3	4	5
BDA-DC										
Sensing	●	●	●	●	●	●	●	●	●	●
Seizing	●	●	●	●	●	•	•	•	•	•
Transforming	●	●	●	●	●	•	•	•	•	•
Environment										
Dynamism	●	●	●	●	●	•	•	•	•	•
Hostility	●	●	●	●	●	•	•	•	•	•
Organizational factors										
Large firms	⊗	●	⊗	●	●	⊗	⊗	●	⊗	●
Small firms	●	⊗	●	⊗	⊗	●	●	⊗	●	⊗
Product Industry	⊗	⊗	●	●	⊗	●	⊗	⊗	●	●
Service Industry	●	●	⊗	⊗	●	⊗	●	●	⊗	⊗
Low cost strategy	●	●	⊗	⊗	⊗	●	●	●	⊗	⊗
Differentiation strategy	⊗	⊗	●	●	●	⊗	⊗	⊗	●	●
Consistency	0.995	0.944	0.960	0.934	0.950	0.935	0.967	0.943	0.966	0.959
Raw coverage	0.077	0.143	0.559	0.501	0.100	0.098	0.074	0.145	0.055	0.051
Unique coverage	0.077	0.143	0.559	0.501	0.100	0.098	0.074	0.145	0.055	0.051
Overall solution consistency	0.955					0.950				
Overall solution coverage	0.430					0.425				

Explorative innovation

The five solutions of high levels of innovation of exploration depict large black circles for BDA-DC in sensing, seizing, and transforming under higher environmental uncertainty of dynamism and hostility. This result means that for firms to achieve innovation of exploration, the high presence of big data & analytics is essential, and the absence of any aspects of BDA-DC can restrict radical innovation under a higher uncertainty environment.

In more detail, specific analysis of solutions 3 and 4 lead to high levels of innovation of exploration, indicating to firms that belong to the small and large size-class and products industry that differentiation competitive strategy. Thus, BDA-DC in sensing, seizing, and transforming combined can ability firms to detect emerging opportunities, develop novel products, and reconfigure business processes through targeted data-generated insight is argued to create radical innovation. These two solutions show high levels of consistency (consistency above 0.9) and explain a substantial number of cases (coverage above 0.5).

Thus, these results demonstrate that BDA-DC in radical innovation under an uncertain environment to create new products and services to meet the needs of emerging customers or markets, satisfying the needs of potential customers and markets through creating and developing new markets and channels.

Moreover, solution 5 shows that organizations in a large service industry with a competitive differentiation strategy also develop radical innovation. The coverage indicates a small (0.100) group of firms with the combination in the sample.

Solutions 1 and 2 demonstrate high levels of consistency (>0.94) and explain a satisfactory coverage (< 0.17) for the large and small size firm in the service industry with the low-cost competitive strategy to create radical innovation by BDA-DC. The dynamism and hostility factors can be imperative for these kinds of organizations. To survive in the uncertain environment, they must reconfigure their business model and process to create radical innovation in a low-cost competitive strategy. Then, BDA-DC in innovation can transform the organization to create business value for the market and customer through explorative innovation to gain a competitive advantage.

Exploitative innovation

The five solutions of high levels of exploitative innovation depict big black circles for BDA-DC in sensing and medium black circles of BDA-DC in seizing and transforming and under medium level degree of environmental uncertainty through dynamism and hostility. Exploitative innovation demonstrates high overall solution consistency (0.95) and adequate overall solution coverage (0.42).

Solution 3 indicates a critical path to high incremental innovation for large firms that compete in the service industry with low-cost competitive strategy. Regarding the environment, firms that operate in moderate competitive conditions are characterized by a medium degree of dynamism and hostility. Solution 3 demonstrates high levels of consistency (0.94) and satisfactory level of coverage (0.145).

Moreover, solution 1 shows an important path to high exploitative innovation for small and product industry with low-cost competitive strategy. However, solution demonstrated high levels of consistency (0.93) and coverage is low, indicating a small portion of sample contemplate this configuration for incremental innovation. The conditions of fairly consistent in solutions 2, 4, and 5 (>0.90) and have sufficient coverage (<0.07), indicating a small portion of the organizations.

Thus, organizations create high levels of incremental innovation through BDA-DC in sensing through analyzing the external factors of customers' needs, and market requirements. However, BDA-DC in seizing and transforming have a medium role in incremental innovation under medium degree of environmental uncertainty for all characteristics of firms (size, industry, and competitive strategy).

DISCUSSION

Theoretical implications

The results of this empirical study extend knowledge of the existing literature in several ways. Findings showcase the BDA value enables dynamic capabilities to create incremental and radical innovation.

To date, there are a few studies of BDA effects on dynamic capabilities in sensing, seizing, and transforming to create incremental and radical innovation (Wamba *et al.*, 2020). Moreover, when BDA-DC in innovation associated with relevant contextual factors (Mikalef and Krogstie, 2020). The different configurations of big data analytics through dynamic capabilities in sensing, seizing, transforming, and contextual factors for achieving higher levels of incremental and radical also indicated certain commonalities and some striking differences.

An interesting point in the analysis is when an organization creates radical innovation through big data analytics. There is a stronger emphasis on dynamic capabilities in sensing, seizing, transforming, and high levels of uncertainty environment, and other firm characteristics are found to be core contributors. Other example, the role of emphasizing sense dynamic capabilities more than seizing and transforming capabilities were found throughout all solutions for exploitative innovation. Hence, the result of exploration and exploitation of innovation are consistent with current empirical studies of big data analytics to enable organizations capabilities to identify, develop, co-develop, and assess technological opportunities that meet customer needs and business opportunities (Mikalef and Krogstie, 2020; Wamba *et al.*, 2020).

The difference between the low-cost and differential competitive strategy regarding the role of creating products and services by incremental and radical innovation is also noted in the current strategic management literature (Porter, 1998). According to the body of competitive strategy and information systems strategies literature, firms should develop organizational capabilities to sense, seize and transform through business processes to create exploration innovation (Yoshikuni and Lucas, 2022) is more present in the competitive strategy of differential, and low-cost strategy to create capacity to gain business process performance to gain productive. Thus, this empirical study contributes to extend the knowledge of strategic management and big data analytics literature.

Practical implications

As stated, the results suggest that managers should create different strategies associated with the BDA initiatives to achieve exploitative and explorative innovation. Furthermore, organizations should also build organizational capabilities under the challenge of external factors. Specifically, the results indicate that radical innovation exists in the service and product industries that build dynamic capabilities in sensing, seizing, and transforming through big data analytics under higher levels of dynamism and hostility. Nevertheless, firms that create exploitative innovation must build dynamic capabilities in sensing to identify opportunities and threats, market requirements, and customer needs under a moderate degree of uncertainty environment to increment existing products and services.

According to current empirical studies of BDA capabilities (Conboy *et al.*, 2020; Olszak and Zurada, 2020; Kristoffersen *et al.*, 2021), organizations adopt a standard strategy to invest in BDA initiatives. However, it could be a principal reason of organizations fails in the BDA initiatives to realize innovative business value. Thus, organizations and managers need to analyze some aspects and they should be prioritized to create innovation.

References

Albors-Garrigos, J., Igartua, J. I. and Peiro, A. (2018) ‘Innovation management techniques and tools: Its impact on firm innovation performance’, *International Journal of Innovation Management*, 22(6). doi: 10.1142/S1363919618500512.

Almeida, M. C. *et al.* (2022) ‘Do Leadership Styles Influence Employee Information Systems Security Intention ? A Study of the Banking Industry’, *Global Journal of Flexible Systems Management*, ahead of p. doi: 10.1007/s40171-022-00320-1.

Amit, R. and Schoemaker, P. J. H. (1993) ‘Strategic Assets and Organizational Rent’, *Strategic Management Journal*, 14(1), pp. 33–46.

Barney, J. B. (1991) ‘Firm Resources and Sustained Competitive Advantage’, *Journal of Management*, pp. 99–120. doi: 10.1177/014920639101700108.

Barney, J. B. (2001) ‘Resource-based theories of competitive advantage: A ten-year retrospective on the resource-based view’, *Journal of Management*, 27(6), pp. 643–650. doi: 10.1016/S0149-2063(01)00115-5.

Benner, M. J. and Tushman, M. L. (2015) ‘Reflections on the 2013 Decade Award--"Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited" Ten Years Later’, *Academy of Management Review*, 40(4), pp. 497–514. doi: 10.5465/amr.2015.0042.

Bharadwaj, A. (2000) ‘A Resource-Based Perspective on Information Technology Capability and Firm Performance’, *MIS Quarterly*, 24(1), pp. 169–196.

Bido, D. S. and Silva, D. (2019) ‘SmartPLS 3: specification, estimation, evaluation and reporting’, *RAEP*, 20(2), pp. 1–31. doi: 10.13058/raep.2019.v20n2.1545.

Blank, T. H. and Naveh, E. (2019) ‘Competition and Complementation of Exploration and Exploitation and the Achievement of Radical Innovation: The Moderating Effect of Learning Behavior and Promotion Focus’, *IEEE Transactions on Engineering Management*. IEEE, 66(4), pp. 598–612. doi: 10.1109/TEM.2018.2849319.

Braganza, A. *et al.* (2017) ‘Resource management in big data initiatives: Processes and dynamic capabilities’, *Journal of Business Research*. The Authors, 70, pp. 328–337. doi: 10.1016/j.jbusres.2016.08.006.

Ciampi, F. *et al.* (2021) ‘Exploring the impact of big data analytics capabilities on business model innovation: The mediating role of entrepreneurial orientation’, *Journal of Business Research*. Elsevier Inc., 123(September 2020), pp. 1–13. doi: 10.1016/j.jbusres.2020.09.023.

Conboy, K. *et al.* (2020) ‘Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda’, *European Journal of Operational Research*. Elsevier B.V., 281(3), pp. 656–672. doi: 10.1016/j.ejor.2019.06.051.

Davenport, T. H., Harris, J. G. and Morison, R. (2010) *Analytics at Work: Smarter Decisions, Better Results*. Boston, Massachusetts, USA: Harvard Business Press.

Donaldson, L. (2014) *The Contingency Theory of Organizations*. Online Pub. SAGE Publications, Inc. doi: http://dx.doi.org/10.4135/9781452229249.

Dong, J. Q. and Yang, C. H. (2018) 'Business value of big data analytics: A systems-theoretic approach and empirical test', *Information and Management*, (October 2017). doi: 10.1016/j.im.2018.11.001.

Dwivedi, P. *et al.* (2023) 'Role of Digitalization in Business Process to Reduce Manpower : A Case of Human Resource Management Process', *International Journal of Economics and Business Administrationt*, XI(1), pp. 66–76.

Eisenhardt, K. M. and Martin, J. A. (2000) 'Dynamic capabilities: what are they?', *Strategic Management Journal*, 21(10–11), p. Pages 1105–1121. doi: 10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E.

Fiss, P. C. (2011) 'Building better causal theories: A fuzzy set approach to typologies in organization research', *Academy of Management Journal*, 54(2), pp. 393–420.

Hair, J. F. *et al.* (2017) *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. 2° Ed. Thousand Oaks: Sage Publications, Inc.

Helfat, C. E., Kaul, A. and Ketchen, D. J. (2023) 'Strategic Management Journal - 2023 - Helfat - Renewing the resource-based view New contexts new concepts and new.pdf', *Strategic Management Journal*, (Special Issue Article), pp. 1–34.

Helfat, C.E. and Raubitschek, R. S. (2018) 'Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems', *Research Policy*. Elsevier, 47(8), pp. 1391–1399. doi: 10.1016/j.respol.2018.01.019.

Helfat, Constance E. and Raubitschek, R. S. (2018) 'Dynamic and integrative capabilities for profiting from innovation in digital platform-based ecosystems', *Research Policy*, 47(8), pp. 1391–1399. doi: 10.1016/j.respol.2018.01.019.

Helfat, C. E. and Winter, S. G. (2011) 'Research prospectives untangling dynamic and operational capabilities: strategy for (n)ever-changing world', *Strategic Management Journal*, 32, pp. 1243–1250. doi: 10.1002/smj.955.

Ilmudeen, A. (2022) *Leveraging IT-enabled dynamic capabilities to shape business process agility and firm innovative capability: moderating role of turbulent environment*, *Review of Managerial Science*. Springer Berlin Heidelberg. doi: 10.1007/s11846-021-00501-9.

Jansen, J. J. P. *et al.* (2006) 'Exploratory Innovation, Exploitative Innovation, and Performance: Effects of Organizational Antecedents and Environmental Moderators', *Management Science*, 52(11), pp. 1661–1674. doi: 10.1287/mnsc.1060.0576.

Jantunen, A. *et al.* (2018) 'Dynamic capabilities, operational changes, and performance outcomes in the media industry', *Journal of Business Research*. Elsevier, 89(February), pp. 251–257. doi: 10.1016/j.jbusres.2018.01.037.

Jantunen, Ari *et al.* (2018) 'Dynamic capabilities, operational changes, and performance outcomes in the media industry', *Journal of Business Research*. Elsevier, 89(June 2017), pp. 251–257. doi: 10.1016/j.jbusres.2018.01.037.

Kristoffersen, E. *et al.* (2021) 'The Effects of Business Analytics Capability on Circular Economy Implementation, Resource Orchestration, Capability and Firm Performance', *International Journal of Production Economics*. Elsevier B.V., 239(April), p. 108205. doi: 10.1016/j.ijpe.2021.108205.

Lewis, B. R., Templeton, G. F. and Byrd, T. A. (2005) 'A methodology for construct development in MIS research', *European Journal of Information Systems*, 14(4), pp. 388–400. doi: 10.1057/palgrave.ejis.3000552.

Li, T. and Chan, Y. E. (2019) 'Dynamic information technology capability: Concept definition and framework development', *Journal of Strategic Information Systems*. Elsevier, 28(4), p. 101575. doi: 10.1016/j.jsis.2019.101575.

Lu, Y. and Ramamurthy, K. (2011) 'Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination', *MIS Quarterly*, 35(4), pp. 931–954.

MacKenzie, S. B., Podsakoff, P. M. and Podsakoff, N. P. (2011) 'Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques', *MIS Quarterly: Management Information Systems*, 35(2), pp. 293–334. doi: 10.2307/23044045.

Melville, N., Kraemer, K. and Gurbaxani, V. (2004) 'Review: Information Technology and Organizational Performance':, *MIS Quarterly*, 28(2), pp. 283–322.

Mendel, J. M. and Korjani, M. M. (2012) 'Charles Ragin ' s Fuzzy Set Qualitative Comparative Analysis (fsQCA) used for linguistic summarizations', *Information Sciences*. Elsevier Inc., 202, pp. 1–23. doi: 10.1016/j.ins.2012.02.039.

Mikalef, P. *et al.* (2020a) 'Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities', *Information and Management*. Elsevier, 57(2), p. 103169. doi: 10.1016/j.im.2019.05.004.

Mikalef, P. *et al.* (2020b) 'Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities', *Information and Management*. Elsevier B.V., 57(2), p. 103169. doi: 10.1016/j.im.2019.05.004.

Mikalef, P. and Krogstie, J. (2018) 'Big data analytics as an enabler of process innovation capabilities: A configurational approach', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11080 LNCS(January), pp. 426–441. doi: 10.1007/978-3-319-98648-7_25.

Mikalef, P. and Krogstie, J. (2020) 'Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities', *European Journal of Information Systems*. Taylor & Francis, 29(3), pp. 260–287. doi: 10.1080/0960085X.2020.1740618.

Mikalef, P. and Pateli, A. (2017) 'Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA', *Journal of*

Business Research. Elsevier B.V., 70, pp. 1–16. doi: 10.1016/j.jbusres.2016.09.004.

Mintzberg, H., Ahlstrand, B. W. and Lampel, J. (2009) *Strategy safari: The complete guide through the wilds of strategic management*. 2nd edn. Harlow:UK: Financial Times Prentice Hall.

Newkirk, H. E. and Lederer, A. L. (2006) 'The effectiveness of strategic information systems planning under environmental uncertainty', *Information & Management*, 43(4), pp. 481–501. doi: 10.1016/j.im.2005.12.001.

Olszak, C. M. and Zurada, J. (2020) 'Big Data in Capturing Business Value', *Information Systems Management*. Taylor & Francis, 37(3), pp. 240–254. doi: 10.1080/10580530.2020.1696551.

Pappas, I. O. and Woodside, A. G. (2021) 'Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing', *International Journal of Information Management*. Elsevier Ltd, 58(February), p. 102310. doi: 10.1016/j.ijinfomgt.2021.102310.

Popovič, A. *et al.* (2018) 'The impact of big data analytics on firms' high value business performance', *Information Systems Frontiers*, 20(2), pp. 209–222. doi: 10.1007/s10796-016-9720-4.

Porter, M. E. (1998) *Competitive advantage: Creating and sustaining superior performance: with a introduction*. Free press. New York, NY: Simon and Schuster.

Ragin, C. C. . (2008) *Redesigning social inquiry: Fuzzy sets and beyond.* Chicago: University of Chicago Press.

Ravichandran, T. (2018) 'Exploring the relationships between IT competence, innovation capacity and organizational agility', *Journal of Strategic Information Systems*. Elsevier B.V., 27(1), pp. 22–42. doi: 10.1016/j.jsis.2017.07.002.

Steininger, D. M. *et al.* (2022) 'Dynamic Capabilities in Information Systems Research: A Critical Review, Synthesis of Current Knowledge, and Recommendations for Future Research', *Journal of the Association for Information Systems*, 23(2), pp. 447–490. doi: 10.17705/1jais.00736.

Teece, D. J. (2007) 'Explicating Dynamic Capabilities: The Nature and Microfoundations of (Suvstainabile) Enterprise Performance', *Strategic Management Journal*, 28(13), pp. 1319–1350.

Teece, D. J. (2018) 'Dynamic capabilities as (workable) management systems theory', *Journal of Management & Organization*, (May), pp. 1–10. doi: 10.1017/jmo.2017.75.

Teece, D. J., Peteraf, M. and Leih, S. (2016) 'Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy', *California Management Review*, 58(4), pp. 13–35. doi: 10.1525/cmr.2016.58.4.13.

Tóth, Z. *et al.* (2015) 'Understanding configurations of relational attractiveness of the customer firm using fuzzy set QCA', *Journal of Business Research Understanding*, 68, pp. 723–734. doi: 10.1016/j.jbusres.2014.07.010.

Wamba, S. F. *et al.* (2020) 'The performance effects of big data analytics and supply chain ambidexterity: The moderating effect of environmental dynamism', *International Journal of Production Economics*, 222(September 2019). doi: 10.1016/j.ijpe.2019.09.019.

Van de Wetering, R., Mikalef, P. and Krogstie, J. (2019) 'Big data is power: Business value from a process oriented analytics capability', *Lecture Notes in Business Information Processing*, 339(August), pp. 468–480. doi: 10.1007/978-3-030-04849-5_41.

Winter, S. G. (2003) 'Understanding dynamic capabilities', *Strategic Management Journal*, 24(10 SPEC ISS.), pp. 991–995. doi: 10.1002/smj.318.

Wolf, C. and Floyd, S. W. (2017) 'Strategic Planning Research: Toward a Theory-Driven Agenda', *Journal of Management*, 43(6), pp. 1754–1788. doi: 10.1177/0149206313478185.

Woodside, A. G. (2014) 'Embrace erform model : Complexity theory , contrarian case analysis , and multiple realities', *Journal of Business Research*. Elsevier Inc., 67(12), pp. 2495–2503. doi: 10.1016/j.jbusres.2014.07.006.

Xie, X. *et al.* (2020) 'Collaborative ties and ambidextrous innovation: insights from internal and external knowledge acquisition', *Industry and Innovation*, 27(3), pp. 285–310. doi: 10.1080/13662716.2019.1633909.

Yayla, A. A. and Hu, Q. (2012) 'The impact of IT-business strategic alignment on firm performance in a developing country setting: exploring moderating roles of environmental uncertainty and strategic orientation', *European Journal of Information Systems*. Nature Publishing Group, 21(4), pp. 373–387. doi: 10.1057/ejis.2011.52.

Yoshikuni, A. C. (2022) *Working Paper da Indústria Alpha - Estudo de caso prático do sistema transacional*.

Yoshikuni, A. C. *et al.* (2023) 'Role of Emerging Technologies in Accounting Information Systems for Achieving Strategic Flexibility through Decision-Making Performance: An Exploratory Study Based on North American and South American Firms', *Global Journal of Flexible Systems Management*. doi: 10.1007/s40171-022-00334-9.

Yoshikuni, A. C. and Dwivedi, R. (2022) 'Big data & business analytics-enabled dynamic capabilities to leverage innovation: an empirical study of developing and validanting scale measure', in Institute, D. S. (ed.) *The 51st Annual Meeting of the Southeast Decision Sciences Institute*. Jacksonville, FL, US, pp. 40–53.

Yoshikuni, A. C. and Dwivedi, R. (2023) 'The role of enterprise information systems strategies enabled strategy-making on organizational innovativeness: a resource orchestration perspective', *Journal of Enterprise Information Management*, 36(1), pp. 172–196. doi: 10.1108/jeim-10-2021-0442.

Yoshikuni, A. C. and Lucas, E. C. (2022) 'The Effect of IS-Innovation Strategy Alignment on Corporate Performance : Investigating the Role of Environmental Uncertainty by Heterogeneity', *International Journal of Innovation and Technology Management*, 2250026(5), p. 28. doi: 10.1142/S0219877022500262.

Zhang, K. *et al.* (2023) ‘Effects of social media usage on exploratory innovation, exploitative innovation and organizational agility: the moderating role of learning goal orientation’, *Internet Research*. doi: 10.1108/INTR-07-2021-0503.

APPENDIX I

Factor loadings (bolded) and cross-loadings

Items	SENS	SEIZ	TRAN	DYNA	HOST	EXPR	EXPL
SENS1	**0,814**	0,592	0,560	0,419	0,471	0,567	0,532
SENS2	**0,788**	0,585	0,581	0,433	0,468	0,482	0,484
SENS3	**0,804**	0,668	0,643	0,633	0,541	0,660	0,578
SENS4	**0,806**	0,575	0,624	0,486	0,455	0,559	0,500
SEIZ1	0,537	**0,805**	0,598	0,412	0,415	0,474	0,418
SEIZ2	0,662	**0,842**	0,596	0,482	0,435	0,539	0,491
SEIZ3	0,577	**0,757**	0,649	0,498	0,378	0,548	0,385
SEIZ4	0,599	**0,744**	0,553	0,535	0,459	0,482	0,409
TRAN1	0,563	0,509	**0,749**	0,531	0,483	0,541	0,479
TRAN2	0,590	0,601	**0,733**	0,509	0,486	0,573	0,464
TRAN3	0,609	0,604	**0,794**	0,557	0,489	0,593	0,433
TRAN4	0,575	0,650	**0,835**	0,511	0,441	0,531	0,439
DYNA1	0,384	0,441	0,483	**0,756**	0,355	0,485	0,395
DYNA2	0,453	0,439	0,480	**0,719**	0,426	0,515	0,475
DYNA4	0,559	0,510	0,572	**0,800**	0,510	0,568	0,501
HOST1	0,472	0,399	0,425	0,443	**0,836**	0,431	0,508
HOST2	0,454	0,409	0,480	0,413	**0,726**	0,424	0,461
HOST3	0,413	0,418	0,433	0,368	**0,627**	0,291	0,337
HOST4	0,462	0,380	0,483	0,479	**0,779**	0,466	0,522
EXPR2	0,504	0,489	0,500	0,483	0,370	**0,753**	0,534
EXPR3	0,453	0,458	0,541	0,474	0,326	**0,679**	0,417
EXPR4	0,583	0,516	0,597	0,583	0,518	**0,809**	0,678
EXPR6	0,617	0,519	0,558	0,559	0,431	**0,803**	0,560
EXPP1	0,480	0,403	0,439	0,497	0,452	0,543	**0,807**
EXPP3	0,495	0,405	0,452	0,412	0,526	0,570	**0,709**
EXPP5	0,533	0,444	0,449	0,513	0,453	0,597	**0,769**
EXPP7	0,528	0,433	0,476	0,465	0,497	0,550	**0,828**

APPENDIX II
Measurement items for constructs
Big data analytics-enabled dynamic capabilities
Sensing dynamic capabilities
Please rate how well the company

[SENS1] Scanning trends in the external environment (such as social-cultural, the federal government, demographic, political, energy, technology, etc.) and identifying new business opportunities.
[SENS2] Identifying changes in the organization's target market.
[SENS3] Identifying new business opportunities in the micro-sector environment (such as suppliers, intermediary customers, state and municipal government, regulatory agency, etc.)
[SENS4] Identifying changes in customer needs.

Seizing dynamic capabilities

[SEIZ1] Developing effective routines for creating potential business solutions to deal with opportunities or threats detected.
[SEIZ3] Developing potential business solutions to meet trends in the external environment to deal with opportunities and threats detected.
[SEIZ5] Developing potential business solutions to meet the organization's target market.
[SEIZ6] Develop new ways of doing business to meet customer needs.

Transforming dynamic capabilities

[TRAN1] Adjust your business processes in response to changes in your business priorities.
[TRAN2] Optimize the use of existing productive resources in new areas with new purposes.
[TRAN3] Integrate new know-how with the company's existing knowledge.
[TRAN4] Develop new business processes to achieve the organization's goals and objectives.

Innovation
Exploratory innovation

[EXPR1] Our organization accepts demands that go beyond existing products and services.
[EXPR2] We invent new products and services.
[EXPR3] We experiment with new products and services in our local market.
[EXPR4] We commercialize products and services that are completely new to our organization.
[EXPR5] We frequently utilize new opportunities in new markets.
[EXPR6] Our organization regularly uses new distribution channels.
[EXPR7] We regularly search for and approach new clients in new markets.

Exploitative innovation

[EXPP1] We frequently refine the provision of existing products and services.
[EXPP2] We regularly implement small adaptations to existing products and services.
[EXPP3] We introduce improved, but existing products and services for our local market.
[EXPP4] We improve our provision's efficiency of products and services.
[EXPP5] We increase economies of scales in existing markets.
[EXPP6] Our organization expands services for existing clients.
[EXPP7] Lowering costs of internal processes is an important objective.

Environmental uncertainty
Dynamism

[DYNA1] Products and services in our industry become obsolete very quickly.
[DYNA2] The product/services technologies in our industry change very quickly.
[DYNA3] We can predict what our competitors are going to do next.
[DYNA4] We can predict when our products/services demand changes.

Heterogeneity

[HETE1] In our industry, there is considerable diversity in: customer buying habits.
[HETE2] In our industry, there is considerable diversity in: nature of competition.
[HETE3] In our industry, there is considerable diversity in: product lines.

Hostility

[HOST1] The survival of this organization is currently threatened by scarce supply of labor.
[HOST2] The survival of this organization is currently threatened by scarce supply of materials.
[HOST3] The survival of this organization is currently threatened by tough price competition.
[HOST4] The survival of this organization is currently threatened by tough competition in product/ service quality.
[HOST5] The survival of this organization is currently threatened by tough competition in product/ service differentiation.

DECISION SCIENCES INSTITUTE

Application of priority queuing system in service operations

Xiaofeng Zhao
College of Business, University of Mary Washington
Fredericksburg, VA 22401
xzhao@umw.edu

ABSTRACT

Customer delays have been common in service industries. The long waiting times for service constitute a critical factor influencing service quality. To reduce customer waiting time, many firms have adopted the priority policy to ensure that higher priority customers receive service timely. This research discusses a practical approach to predict the waiting times for customers with priorities. It presents a mathematically tractable expression for waiting time in a general queue, and the results can be implemented in a computer program. Simulation experiments are reported to demonstrate the reliability of the approach.

KEYWORDS: queuing theory, priority queue, service operations, approximation methods

INTRODUCTION

Many manufacturing firms have used various priority policies when dealing with items or jobs. For example, a production scheduler examines the list of items to be produced and then sequences them based on the priority rule. In this research, we apply the priority scheme in service operations since many firms have adopted the priority strategy that categorizes their customers into priority queue classes based on the price (Babad et al. 1996; Kumar 1997; Allon and Federgruen 2008; Jouini and Roubos 2014). Customers with high-priority types receive their services faster than those with low priority (with a discounted price). Most e-business companies, such as Amazon and Ebay, deliver items with different prices and delivery time options (So and Song 1998). The application of priority queues is also a commonplace in the healthcare industry. Many hospitals have adopted the priority policy to ensure that higher-priority patients receive emergency medicine service timely (Davis and Heineke 1998; Ferrand 2018). For instance, the emergency department has various patients with diverse diseases and injuries at different acuity levels. Patients with varying levels of acuity must be classified and treated within an appropriate period (Siddharthan 1996; Cochran and Roche 2009; Hou and Zhao 2020).

This research aims to propose a practical queuing approach to predict the waiting times for customers with different priorities. The queue system is modeled as a general multi-server system with a non-preemptive priority policy. Based on the models for the Markov priority queues, we approximate the waiting time for a general priority queue using the concept of isomorphism. The approximations only require the mean and the standard deviation of the inter-arrival and the service time distributions. The approximations have exact expressions for waiting time with Markov queues and reduce the computation complexity without iterations. The research results can be implemented in a computer program to conduct what-if analysis with different input parameters. Simulation experiments show the results have reasonable accuracy for the service operations.

The rest of the paper is organized as follows. The Markov priority queuing models are analyzed in section 2. In section 3, approximation models are developed using the concept of isomorphism. Section 4 reports the comparisons of the approximations and the simulation experiments. Section 5 concludes the paper.

ANALYTICAL MODELS

In this research, we consider the standard steady state $GI/G/n$ queuing system with unlimited waiting room, the first come first served discipline and independent sequence of independent and identically distributed (i.i.d) inter-arrival times and service times. The following closed-form model for the mean waiting time in the GI/G/n queue is widely used in queuing theory literature (Shore 2006; Hopp and Spearman 2016; Zhao and Hou 2014).

$$W_q(GI/G/n) = \left(\frac{c_a^2 + c_s^2}{2}\right)\left(\frac{\rho^{\sqrt{2(n+1)}-1}}{n(1-\rho)}\right)\left(\frac{1}{\mu}\right) \qquad (1)$$

c_a is the coefficient of variation of inter-arrival time and c_s the coefficient of variation of service time. The model has several advantages (Whitt 2004). For instance, it does not require any iterative algorithm to solve and can be easily implemented into a computer program. Coupling the single-station approximation with the multiple-server is possible for analyzing the queuing network performance. Our research uses the formula when calculating mean customer waiting time for GI/G/n queue without priority.

Markov priority queue

The model above follows the first come first served discipline. In priority schemes, customers with high priorities are served before those with low priorities. There are two further refinements in priority situations, preemption, and non-preemption. In preemptive cases, a customer with the highest priority can enter service immediately, even if another with lower priority has already been served when the higher customer arrives. A priority discipline is non-preemptive if no interruption applies and the highest-priority customer moves to the head of the queue. The customer cannot receive service until the current customer has completed the service, even though this customer has a lower priority. We only study a non-preemptive approach in this paper as it is preferred in most service applications.

The priority queue has been studied extensively in queuing literature (Tabet 1992; Xie et al. 2008; Zhang and Shi 2010; He 2012; Yoon 2018; Hou and Zhao 2020). In this paper, we consider a queuing system consisting of n identical servers that serve N types of customers: type 1, type 2. . . and type N customers, that are queue 1, 2…and N, respectively. Type N customers have the highest service priority, type N-1 the second highest service priority . . . and type 1 the lowest service priority. When a server is available, a customer from the non-empty queue of the highest priority is selected. If some servers are serving type j customers when a type k customer arrives, for j < k, there is no idle server, and type j customers are the lowest priority customers in service, then one of the type j customers in service is pushed back to queue j and the server serves the type k customer. The type j customer will resume or repeat its service if a server is available to serve type j customers.

For a Markov queuing system, type 1,2,… and N customers arrive according to independent Poisson processes with parameters λ_1, λ_2…and λ_N, respectively. The service times of type 1, 2… and N customers are exponentially distributed with parameters μ_1, μ_2…and μ_N, respectively. The arrival processes and service times are independent. Since the service time of a customer is exponentially distributed, it makes no difference to assume that it's interrupted. The number of priority classes can be any number greater than one. Suppose there can be more than one customer in any given priority class. Within each priority class, the FIFO discipline holds.

We refer to the results of Kleinrock (1976) and Gross and Harris (2002) for non-preemptive Markov systems with many priorities. Suppose that the customers of the kth priority (the smaller the number, the higher the priority) arrive before a single channel according to a Poisson distribution with parameter $\lambda_k\,(k = 1,2,...r)$ and that these customers wait on a FIFO basis within their respective priorities. Let the service distribution for the kth priority be exponential with mean $1/\mu_k$.Whatever the priority of a unit in service, it completes its service before another item is admitted.

We define $\rho_k = \dfrac{\lambda_k}{\mu_k}$ $\quad (1 \le k \le r)$ and $\sigma_k = \sum_{i=1}^{k} \rho_i \quad (\sigma_0 \equiv 0, \sigma_r \equiv \rho)$

The system is stationary for $\sigma_r = \rho = \sum_{k=1}^{r} \rho_k < 1$. we have $W_q^{(i)} = \dfrac{\sum_{k=1}^{r} (\rho_k / \mu_k)}{(1-\sigma_{i-1})(1-\sigma_i)}$.

The calculation for the multiple-server case is similar to that of the proceeding analysis except that service is governed by identical exponential distributions for each priority at each of s channels. We assume no service time distinction between priorities for multiple channels. Otherwise, the mathematics becomes quite intractable.

We define $\rho_k = \dfrac{\lambda_k}{s\mu_k}$ $\quad (1 \le k \le r)$ and $\sigma_k = \sum_{i=1}^{k} \rho_i \quad (\sigma_r \equiv \rho = \lambda / c\mu)$

From Kleinrock (1976) and Hillier and Lieberman (1986), we have

$$W_q = \frac{1}{A \cdot B_{k-1} \cdot B_k} + \frac{1}{\mu} \qquad for \;\; k = 1,2,......N$$

Where $A = s!(\dfrac{s\mu - \lambda}{\rho^s}) \sum_{j=0}^{s-1} \dfrac{\rho^j}{j!} + s\mu$, $B_0 = 1$, $B_k = 1 - \dfrac{\sum_{i=1}^{k} \lambda_i}{s\mu}$ $\qquad for \;\; k = 1,2,......N$

λ_i: mean arrival rate for priority class i, for i=1, 2, .N , $\lambda = \sum_{i=1}^{N} \lambda_i$.

Little's formula applies to individual priority class, so

$L_k = \lambda_k W_k \qquad for\; k = 1,2,.......N$. Hence, $L_q = \dfrac{\lambda_k}{A \cdot B_{k-1} \cdot B_k}$.

$$L_{q1}=\frac{\lambda_1}{A\cdot B_0\cdot B_1}=\frac{\lambda_1}{\left[s!(\frac{s\mu-\lambda}{(\lambda/\mu)^s})\cdot\frac{\lambda}{\mu}+s\mu\right]\cdot(1-\frac{\lambda_1}{\mu})}$$

$$=\frac{\lambda_1}{\left(\frac{\mu^2}{\lambda}\right)(1-\frac{\lambda_1}{\mu})}=\frac{\lambda^2}{\mu(\mu-\lambda)}\frac{\lambda_1}{\lambda}\left(1-\frac{\lambda}{\mu}\right)\Big/(1-\frac{\lambda_1}{\lambda}\frac{\lambda}{\mu})$$

$$=\frac{L_q\cdot Fract1\cdot(1-\rho)}{(1-Fract1\cdot\rho)}.$$

$$L_{q2}=\frac{\lambda_2}{A\cdot B_1\cdot B_2}=\frac{\lambda_2}{\left(\frac{\mu^2}{\lambda}\right)\cdot(1-\frac{\lambda_1}{\mu}-\frac{\lambda_2}{\mu})\cdot(1-\frac{\lambda_1}{\mu})}$$

$$=\frac{\lambda\cdot\frac{\lambda_2}{\mu^2}}{(1-\frac{\lambda_1}{\mu})((1-\frac{\lambda_1}{\mu}-\frac{\lambda_2}{\mu})}$$

$$=\frac{\frac{\lambda^2}{\mu(\mu-\lambda)}\frac{\lambda_2}{\lambda}\left(1-\frac{\lambda}{\mu}\right)}{(1-\frac{\lambda_1}{\lambda}\frac{\lambda}{\mu})((1-\frac{\lambda_1}{\lambda}\frac{\lambda}{\mu}-\frac{\lambda_2}{\lambda}\frac{\lambda}{\mu})}$$

$$=\frac{L_q\cdot Fract2\cdot(1-\rho)}{(1-\frac{\lambda_1}{\lambda}\rho)((1-\frac{\lambda_1}{\lambda}\rho-\frac{\lambda_2}{\lambda}\rho)}$$

$$=\frac{L_q\cdot Fract2\cdot(1-\rho)}{(1-Fract1\cdot\rho)(1-Fract1\cdot\rho-Fract2\cdot\rho)}.$$

Similarly, we can derive L_{q3} and L_{q4} etc.

For multi-servers, $n\neq 1$

$$A=s!(\frac{s\mu-\lambda}{\rho^s})\sum_{j=0}^{s-1}\frac{\rho^j}{j!}+s\mu \qquad \text{(Note here } \rho=\frac{\lambda}{\mu}\text{)}$$

$$B_0=1;\ B_k=1-\frac{\sum_{i=1}^{k}\lambda_i}{s\mu}. \qquad for\quad k=1,2,......N$$

We use the same reasoning: $L_q=[\frac{(\lambda/\mu)^s\lambda\mu}{(s-1)!(s\mu-\lambda)^2}]P_0$

$$P_0=\left[\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!}+\frac{(\lambda/\mu)^s}{s!}\frac{s\mu}{(s\mu-\lambda)}\right]^{-1}$$

$$L_{q1} = \frac{\lambda_1}{A \cdot B_0 \cdot B_1}$$

$$= \frac{\lambda_1}{\left[s!(\frac{s\mu - \lambda}{(\lambda/\mu)^s}) \cdot \frac{(\lambda/\mu)^0}{1!} + s\mu \right] \cdot 1 \cdot \left[1 - \frac{\lambda_1}{s\mu} \right]}$$

$$= \frac{L_q \cdot \frac{\lambda_1}{\mu}(1 - \frac{\lambda}{s\mu})}{(1 - \frac{\lambda_1}{\lambda}\frac{\lambda}{s\mu})}$$

$$= \frac{L_q \cdot Fract1 \cdot (1 - \frac{\lambda}{s\mu})}{(1 - Fract1 \cdot \frac{\lambda}{s\mu})}.$$

Similarly, we can have derive L_{q2} , L_{q3} , L_{q4} etc. By using Little's rule, we can obtain waiting times $W_{q1}, W_{q2,} W_{q3,}$ *and* W_{q4} *etc.*

Approximation for the general queue with priority

In practice, the arrival process may not be Poisson distribution, and service time distribution is not always exponential distribution (Tabet 1992; Lin et al. 2004). It is necessary to consider different probability distributions other than Poisson and exponential distributions for arrival and service processes (Kim and Kim 2015; Hanbali et al. 2015). In this section, we consider the general GI/G/n queuing system with *n* homogeneous servers in parallel, and the inter-arrival times and the service times are independent and identically distributed (i.i.d.) for each priority class. We focus on the non-preemptive $GI/G/n$ system with many priorities. Due to the difficulty of the determination of stationary priorities of $GI/G/n$, the calculation of stationary probabilities in a non-preemptive general queuing system with more than two priority classes has no exact results. However, the priority queuing models should not be oversimplified merely to permit calculation results. We need careful analysis of priorities in practice when considering the waiting time with priority customers.

Due to generality in the stochastic structures in the GI/G/n queues, there is no exact mathematical formula available for mean waiting time with different classes. We focus on the GI/G/s system with the assumption that the general inter-arrival time and service time distributions are specified by their first two moments only. We use the similar approximation approach analogous to the $M/M/n$ priority queue to derive the formula for the general priority queue.

Define $W_q(GI/G/n)$ as the mean waiting time of the arbitrary customer in the GI/G/n queue and $W_q(M/M/n)$ the mean waiting time for the M/M/s queue. For the GI/G/n queue without priority, it is well known (Hopp and Spearman 2011; Gross and Harris 2002) that

$$W_q\ (GI/G/n) = (1/2)(c_a^2 + c_s^2)W_q\ (M/M/n) \qquad (2)$$

Formula (2), exact for the M/M/s queue $(c_a^2 = 1, c_s^2 = 1)$, is a two-moment approximation for the mean waiting time in the general queue. It is dependent only on the first two moments of inter-arrival time distribution and service time distribution. Two-moment approximations for queuing characteristics have sufficient accuracy for practical purpose, since extreme cases which need higher moments tend not to arise (Kimura 1986; Bertsimas 1990; Alotaibi and Liu 2013). The approximation offers several advantages and where $(1/2)(c_a^2 + c_s^2)$ refers to variability. Both Whitt (2004) and Hopp and Spearman (2011) discussed the structure of the approximation for one class without priority, which neatly separates into three terms: a dimensionless variability term, a utilization term, and a time term.

We use the isomorphism method to approximate the GI/G/n priority queue which is analogous to the M/M/n priority queue. The concept of isomorphism involves obtaining numerical results for a queueing system that has some points of similarity with a given queue. This concept has been used to approximate waiting time distribution and variance of waiting time (Seelan and Tijms 1984). In the spirit of Seelan and Tijms (1984) and Zhao and Hou (2014), we conjecture that the mean waiting time for each priority class in GI/G/n has the same relations to those of the M/M/n priority queue. Define $W_q^{(i)}(GI/G/n)$ as the mean waiting time for class *i* in the general queue and $W_q^{(i)}(M/M/n)$ the mean waiting time in a Markov queue. Then, for each class in the GI/G/s queue, we obtain

$$W_q^{(i)}(GI/G/n) = (1/2)(c_a^2 + c_s^2) W_q^{(i)}(M/M/n) \qquad (3)$$

Model (3) is used to approximate the mean waiting times in non-preemptive priority queues in the GI/G/n, which are exact for each class in $W_q^{(i)}(M/M/n)$ queues with priority when both c_a, c_s equal to 1. The mean waiting time for different priorities in the M/M/n queue plays a crucial role in the analysis. Formula 3 includes the variability term and represents an intermediate case for the priority queue analogous to that by the worst case. If the variability factor is less than one, then the waiting time for the GI/G/n priority queue will be shorter than those for the M/M/n priority queue. The coefficients of variation c_a and c_s are used to reflect the level of variability in service times and inter-arrival times.

SIMULATION AND NUMERICAL COMPARISIONS

To evaluate the approximations' accuracy, we conduct experiments using the Extend simulation program (www.extendsim.com). The testing of our approximations has been based on extensive simulation experiments. In this simulation experiment, we performed independent replications and estimated 95 % confidence intervals. Gamma distributions are used as general distribution. For Gamma distribution, when shape parameter k is positive integer, Gamma is reduced to Erlang. When k=1, it is exponential. When k→∞, it is deterministic.

We noticed that as utilization increases, the standard deviation of the errors increases accordingly. As the number of servers increases, the error standard deviation decreases. There are two standard ways to measure the quality of queuing approximations: absolute difference and relative percentage error (Whitt 2004). We contend that neither procedure alone is usually suitable over the entire range of values. We can obtain satisfactory results if either the absolute difference is below a critical threshold or the relative percentage error is below another critical threshold. Thus, a final adjusted measure of error (AME) might be:

$$Error = \min\{|eaxct - approx.|,\ 100(|exact - approx.|)/exact.\}.$$

Either the relative percentage error or the absolute difference should be small. Here we have simulation results corresponding to different experiments. These tables display expected mean and standard deviation of cycle time in specific queuing systems. The difference and relative error analysis are displayed in a separate spreadsheet. For those cases with both $c_a, c_s \leq 1.25$, the approximations appear to be remarkably accurate.

The approximation (spreadsheet) results are compared to the simulation results. The errors are calculated by using $\%Error = 100|(\text{spread sheet}\ \ \sigma - \text{simulation}\ \ \ \sigma)|/simulation\ \ \sigma$. Different cases with a combinations of parameters are compared. Simulation experiments confirm that the approximations perform reliability well across a wide range of cases. In most of these cases the standard deviation of the time in the system obtained with the spreadsheet was within 10% of that obtained in the simulation. The limitation is the assumption that the coefficients of variation of the inter-arrival times and the service times are between 0 and 1.5, which is usual in practice. As noted by Whitt (2004), greater variability leads to less reliable approximation, as such descriptions are dependent more critically on the missing information.

In most cases, the standard deviation of the time in the system obtained with the spreadsheet was within 10% of that obtained in the simulation. The limitation is that our result is under the assumption that the coefficients of variation of the inter-arrival times and the service times are between 0 and 1.25, which is usual in practice. When coefficients of variation are greater than1.5, the performance of the queue itself becomes very unstable. As noted by Whitt (2004), greater variability means less reliable approximation, because such descriptions depend more on the missing information.

DISCUSSION AND CONCLUSION

We have developed mathematically tractable expressions for the waiting time for Markov priority queues. We provide an approximation for the mean waiting time in the system for a general queue with priority. The measurement requires only the mean and standard deviation or the coefficient of variation of the inter-arrival and service time distributions, and the number of servers. Simulations experiments demonstrate that the approach for the steady-state performance is sufficiently accurate for the waiting time with multiple priority classes. The quality of the approximations is not the same for all cases, but in comparison to the simulations has proven to give good approximations (within ± 10%). The research combines the current research results in queuing theory which provides more applications than Markov queue models. A significant feature of the approximation methods is that the explicit expression requires no iterative algorithm and is implementable in a computer program.

Using priorities increases the variability of waiting times: the higher the percentage of customers getting preferential service, the higher the variability. Because variability adds uncertainty to business outcomes, using priority rules in processing waiting line customers should be carefully considered. It should be limited to only a small percentage of the arrival population if used. The literature (Kim and Kim 2015; Hanbali et al. 2015; Ferrand et al. 2018) shows that some models have been developed to determine the increased variability in average waiting time when using non-preemptive and preemptive priorities. These models help managers determine the degree of reduction in the average waiting time for higher-priority customers and the concomitant increase in waiting time for lower-priority customers.

Because priority rules do not all affect performance measures to the same degree, a manager should select a policy that best addresses the performance measure that is most important for the service operations. Suppose all customers must go through the same sequence of service operations, the queuing analysis models discussed above can be used to determine which priority rule provides the highest performance measures for a specific firm. Admittedly, the approximation model in this paper has some limitations. It is worthwhile to extend the current research by developing a more sophisticated system when the coefficients of variation of arrival and service processed are high.

REFERENCES

Allon, G., & Federgruen, A. (2008). Service competition with general queueing facilities. *Operations Research*, *56*(4), 827-849.

Bertsimas, D. (1990). An analytic approach to a general class of G/G/s queuing systems *Operations Research* , 38(1),139.

Çil, E. B. (2021). Managing service systems via disguised queues: the role of retaliating customers. *Decision Sciences*, *52*(2), 427-454.

Davis, M.M. and Heineke, J. (1998), How disconfirmation, perception and actual waiting times impact customer satisfaction , *International Journal of Service Industry Management*, 9(1), 64-73.

Hopp, W.J. and Spearman, M.L. (2000). *Factory Physics*, 2nd ed. Irwin/McGraw-Hill.

Hou J. and Zhao, X (2020). Using a priority queuing approach to improve emergency department performance, *Journal of Management Analytics*, 7(1), 28-43.

Houston, M. B., Bettencourt, L. A., & Wenger, S. (1998). The relationship between waiting in a service queue and evaluations of service quality: a field theory perspective. *Psychology & Marketing*, *15*(8), 735-753.

Jones, P., & Peppiatt, E. (1996). Managing perceptions of waiting times in service queues. *International Journal of Service Industry Management*, *7*(5), 47-61.

Kimura, T. (1986). A two-moment approximation for the mean waiting time in the GI/G/s queue, *Management Science,* 32, 751-763.

Lin, Y. T., Xia, K. N., & Bei, L. T. (2015). Customer's perceived value of waiting time for service events. *Journal of Consumer Behavior*, *14*(1), 28-40.

Kleinrock, L. (1976). *Queuing Systems*, Volume I & II: Theory. John Wiley and Sons.

Sakasegawa, H. (1977). An approximate formula Lq = αβ ρ / (1 – ρ), *Ann. Inst. Statist. Math*.

29,67-75.

Parasuraman, A., Zeithaml, V.A. and Berry, L.L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4),41-50.

Seelen, L.P. and TIJMS. (1984). Approximations for the conditional waiting times in GI/G/c queue, *Operations Research letters,* 3. 183- 190.

Shore, H. (1988). Simple approximations for the GI/G/c queue-I: The steady-state probabilities *The Journal of the Operational Research Society*, 39(3), 279-284.

Shortle, J. F., Thompson, J. M., Gross, D., & Harris, C. M. (2018). *Fundamentals of Queueing Theory* . John Wiley & Sons.

Taylor, S. (1994). The effects of filled waiting time and service provider control over the delay on evaluations of service. *Journal of the Academy of Marketing Science*, *23*(1), 38-48.

Whitt, W. (2004).A diffusion approximation for the GI/G/n/m queue. *Operations Research,* 52 (6), 922-941.

Zeithaml, V. A., Berry, L. L., & Parasuraman, A. (1993). The nature and determinants of customer expectations of service. *Journal of the Academy of Marketing Science*, *21*(1), 1-12.

Zhao, X., Hou, J., & Gilbert, K. (2014). Measuring the variance of customer waiting time in service operations. *Management Decision*. 52(2), 296-312.

DECISION SCIENCES INSTITUTE

Are the Lean Maturity Models Matured? A Criteria-based Evaluation

Padmaka Mirihagalla
Széchenyi István University
Email: padmaka@gmail.com

Gyula Vastag
Széchenyi István University
Email: vastag.gyula@sze.hu

ABSTRACT

The journey of maturing in lean requires a compass to tell an enterprise its current position and the way ahead. Yet, no Lean maturity model (LMM) has become the commonly accepted yardstick of lean maturity. The paper addresses this gap by evaluating 27 qualified LMMs, revealing their critical weaknesses. In achieving this, the paper introduces the OVRGP framework for evaluating MMs (opportunity, validity, reliability, generalizability, and process integrity) with 17 underlying criteria to be satisfied. 9 critical weaknesses of current LMMs are revealed with in-depth remediations recommended for future developers of LMMs.

KEYWORDS: lean maturity, maturity models, lean integration, measure lean, formative index, measurement model, lean journey

DECISION SCIENCES INSTITUTE
Bitcoin Tweet Sentiment Analysis and Classification

Nathan Burleyson*
nburleys@depaul.edu

Vatsal Parikh*
vparikh3@depaul.edu

Yashel Khan*
ykhan14@depaul.edu

Ali Amin*
aamin12@depaul.edu

Ilyas Ustun*
iustun@depaul.edu

*: DePaul University

ABSTRACT

This paper highlights the application of Natural Language Processing (NLP) techniques in the analysis of Twitter data. With the growing prominence of cryptocurrency in recent years, understanding public sentiment has become pivotal for comprehending price fluctuations. In this study, we used five distinct models to derive sentiment scores from tweets. For the classification task, Long Short-Term Memory (LSTM), a well-suited variant of Recurrent Neural Networks (RNNs), was employed. The results demonstrate the effectiveness of LSTM in achieving high accuracy scores for sentiment analysis. This research underscores the utility of NLP techniques and LSTM models in the analysis of Twitter data for cryptocurrency sentiment assessment.

KEYWORDS: Sentiment Analysis, Bitcoin, Cryptocurrency, NLP

INTRODUCTION

The ascent of Bitcoin has played a pivotal role in the transformation of the economic landscape, ushering in the era of cryptocurrency growth. As a decentralized digital currency, its value is intricately linked to the shifting tides of public sentiment, reflected in the fluctuations driven by buying and selling activities. Consequently, comprehending the prevailing sentiment surrounding Bitcoin is fundamental to gaining insights into the dynamics of its price. Social media platforms like Twitter have emerged as invaluable sources for gauging this sentiment. To this end, we leveraged an extensive Twitter dataset sourced from Kaggle, comprising over 4 million tweets containing the keywords #Bitcoin, #BTC, BTC, or Bitcoin. This dataset served as a robust foundation for deriving sentiment scores, which were subsequently analyzed using Long Short-Term Memory (LSTM). The tweets encompassed the time span from the outset of 2021 to March 2023. Prior to sentiment score computation, essential preprocessing steps were meticulously undertaken.

Our approach encompassed the integration of five distinct sentiment score extraction techniques: SentiWordNet, TextBlob, VADER, AFINN, and a calculation based on positive and negative word counts. This amalgamation facilitated the implementation of a majority voting system, enabling the derivation of a consolidated sentiment score for each individual tweet. This comprehensive sentiment scoring strategy, prior to the LSTM analysis, ensured a more robust and resilient approach. It mitigated the risk of being overly reliant on a single model, which could potentially introduce bias or limitations if used in isolation.

LITERATURE REVIEW

Hasan et al. (2019) focus on the use of Bag of Words and Term Frequency-Inverse Document Frequency as features for an NLP model. They specifically looked at tweets during this time. From a few years prior, Chong et al. (2014) used sentiment lexicons to extract sentiment from tweets. This was an important case of dealing with the informal and unstructured nature of tweet data.

Dang et al. (2020) did a comparative study between deep learning models, to further show the effectiveness of deep learning in handling tweet data. The accuracy achieved in these deep learning experiments echoes with our work with the LSTM Recurrent Neural Network.

More issues NLP techniques face that are not specific to Twitter include distinguishing between sarcastic and non-sarcastic language. Mukherjee and Bala (2017) discuss supervised learning like Naive Bayes to figure out what is sarcastic and what is not. This was under the guise of learning more about customer opinions. Weerasooriya et al. (2016) attempted to extract keywords using Stanford CoreNLP Part-of-Speech tagger to deal with the brevity of tweets. A hybrid approach was used here meshing machine learning and rule-based techniques.

The work of Shukla et al. (2022) goes into detail about the process of sentiment analysis, as well as challenges and future insights. Included are data preprocessing techniques, different sentiment analysis models, and evaluation metrics. Many of these concepts are directly applied in the work of Shanmugavadivel et al. (2022) which aimed to develop a system which performs sentiment analysis and offensive language identification on code-mixed data. Many different models were applied and compared, with Adapter-BERT yielding the best results for both sentiment analysis and offensive language identification.

Further research was done on BERT models. The comparative study of Bello et al. (2023) highlighted the improvements made by using BERT for sentiment analysis of tweets in comparison to more traditional models and techniques. The work of Pota et al. (2021) dealt with performing sentiment analysis on tweets in both English and Italian. They detail a novel and significant preprocessing approach which allows the data to be more proficiently used by their BERT model.

Cheng et al. (2023) explored the GameStop and ‘meme-stock’ phenomenon of early 2021. They used VADER to perform sentiment analysis of the r/WallStreetBets subreddit on the Reddit social media platform and its effects on GameStop share prices. The results suggest that there is a strong relation between positive sentiments on social media and when the market is up, but also that social media has weak influence in affecting the market when the market is going down.

In their paper, Baroiu and Ene (2022) explore the relationship between Twitter sentiment and Bitcoin price in the context of Web 3.0. The relationship between the two is analyzed using a Vector Autoregressive model. The results indicate that there is a mutual influence between Twitter sentiment and Bitcoin price, with Twitter sentiment reacting more quickly to information while Bitcoin price takes longer to respond.

The work of Edgari et al. (2022) investigates the impact of Twitter sentiment analysis on Bitcoin price during the COVID-19 pandemic. They analyze the sentiment of Bitcoin-related tweets using VADER and combine it with Bitcoin price data. The authors develop a predictive model using XGBoost and compare the performance of the model with and without Twitter sentiment analysis. The results indicate that Twitter sentiment has an impact on Bitcoin price.

Khurshid (2021) explores the relationship between media sentiment and the prices of cryptocurrencies, specifically Bitcoin and Cardano. The study utilizes sentiment analysis on various data sources, including tweets, tweet volume, and Google Trends data, to predict price fluctuations. The paper also discusses different machine learning algorithms such as logistic regression and linear regression used for prediction purposes. The author highlights the potential of sentiment analysis combined with machine learning tools to improve the efficiency of cryptocurrency price predictions.

The work of Haritha et al. (2023) proposes an algorithm that combines Twitter sentiment analysis and historical price data to predict the price of Bitcoin. The algorithm uses a BERT-based model to forecast the sentiment of a set of tweets and a Gated Recurrent Unit to forecast the price of Bitcoin. The results show that the sentiment prediction has a MAE of 9.45%, while the price prediction has a MAE of 3.6%.

Tung et al (2023) addresses the need for sentiment analysis of memes to identify and censor harmful content on social media platforms. The paper proposes a multimodal semi-supervised learning approach called SemiMemes, which combines auto-encoder and classification tasks to leverage the large amount of unlabeled meme data available on the internet. The approach utilizes both images and texts, as memes' meanings often come from both modalities. The proposed approach shows promise in effectively analyzing and detecting harmful content in memes, contributing to the field of memes sentiment analysis.

Purohit & Patheja (2023) examines Amazon customer reviews using the bag-of-words model, which was used to extract features from review comments to form feature vectors and assign a score to each word in the grouped review. Then each review was grouped into a positive, negative, or neutral category. To evaluate the performance of the model, the accuracy, precision, recall, F1-score, sensitivity, and specificity were analyzed. It was found that the bag-of-words model reflected high scores for each of the metrics, indicating a suitable model to convey the sentiments of the reviews.

Kim et al (2018) examines a collection of 8-K financial reports by Citigroup Inc., JP Morgan Chase and CO, Wells Fargo and CO, and Goldman Sachs Group Inc. using a Word2vec model in order to embed words from financial reports to a continuous space. Several prediction models were used to present the data; including logistic regression, random forest, multinomial Naïve Bayes, support vector machine, Naïve Bayes SVM, and supervised PV. The sentiments were compared to stock price movement visualizations to conclude if the sentiments in financial reports could be used to predict stock price. It was found that this is indeed the case.

Diana (2023) examines data collected from social media and digital prints in order to calculate an engagement risk priority number, which is used to allow airports to assess community engagement risks and give information to stakeholders. This is found by using the polarity, surprise, and subjectivity of the sentiments from the collected data through BERT, VADER, and Textblob models. The models were also designed to separate opinions that had any detectable sarcasm from non-sarcastic opinions.

Zhang (2023) examines reviews using sentiment methods to analyze polarizing trends in reviews, while detecting spam. This process is done by processing the data on the Tsetlin machine where the data can be decomposed into several patterns. To accomplish this the review data is converted to binary representation and is processed through SVM, Logistic Regression, Decision Trees, and Naive Bayes models.

The work of Bansal and Kumar (2022) examines online hospital reviews using SentiWordNet in order to better understand reviews and provide better information to patients. Sentiments were used to rank hospitals based on their reviews.

Hamed (2023) examines a collection of reddit posts, using text classification and Bi-LSTM to analyze sentiments and detect fake news headlines. It was found that the proposed model demonstrated suitable performance in detecting fake news, given the high accuracy of the model.

DATA

Data Exploration

We are using twitter data which consists of tweets having #Bitcoin and #btc hashtag. Collection was started on 6/2/2021, with an initial 100,000 tweets, and has continued daily. The data totally consists of more than 4 Million records with 13 columns. Columns in the dataset are user_name, user_location, user_description, user_created, user_followers, user_friends, user_favourites, user_verified, date, text, hashtags, source and is_retweet. The explanations of the tweets can be seen in Table 1.

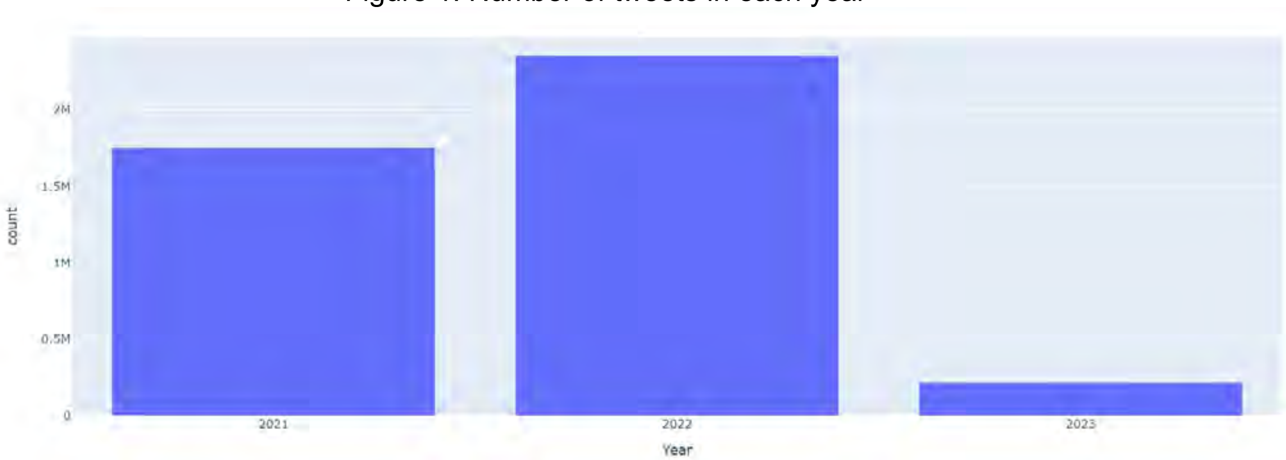
Table 1: Variables present in the data

Columns	Description
user_name	The name of the user, as they've defined it
user_location	The user-defined location for this account's profile
user_description	The user-defined UTF-8 string describing their account
user_created	Time and date, when the account was created
user_followers	The number of followers an account currently has
user_friends	The number of friends an account currently has

user_favourites	The number of favorites an account currently has
user_verified	When true, indicates that the user has a verified account
date	UTC time and date when the Tweet was created
text	The actual UTF-8 text of the Tweet
hashtags	All the other hashtags posted in the tweet
source	Utility used to post the Tweet

Our data consists of bitcoin tweets spanning over three years. We can see in Figure 1 that most of the tweets are from the year 2022 and which is then followed by 2021.

Figure 1: Number of tweets in each year



In Figure 2, it can be seen that wrds like project, blockchain, and buy are the most appearing words in the tweets. Figure 3 shows the word clouds for stop-words and tweets using the first few hundred tweets.

Figure 3: Wordclouds of stop-words and tweets, respectively

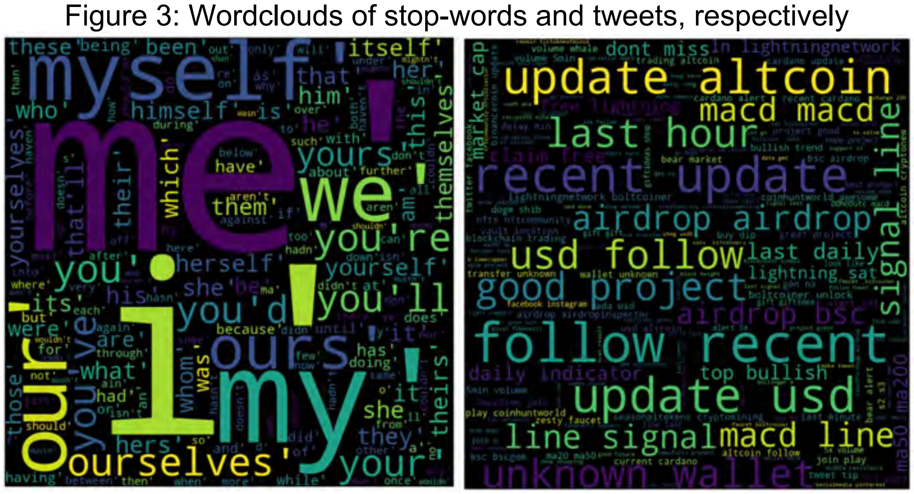
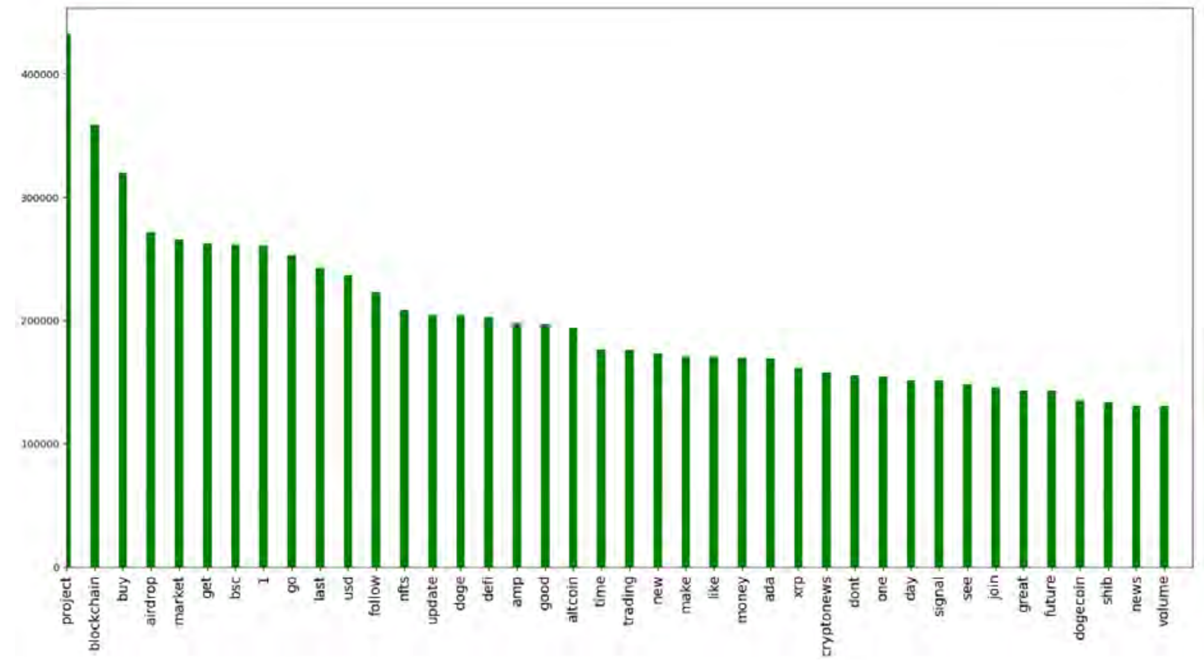


Figure 2: Frequency of words in the preprocessed data



Data preprocessing

The data preprocessing phase encompassed the removal of anomalies and the implementation of fundamental NLP-driven alterations. Notably, we converted the tweet text to lowercase, eliminated punctuation, and removed both common and rare words. A critical decision was made to opt for lemmatization over stemming as the preferred data transformation technique. The rationale behind this choice was the desirability of transforming words into their base forms rather than merely stripping suffixes. This decision paved the way for our subsequent sentiment analysis.

The removal of stopwords and the application of lemmatization play a pivotal role in preparing the tweets for sentiment analysis. Stopwords, such as "and," "the," and "is," are ubiquitous terms in everyday language that contribute little to sentiment analysis, as they are consistently present across various contexts. By eliminating these stopwords, we effectively reduce noise in the dataset, thereby enhancing the precision of sentiment detection. Subsequently, lemmatization is applied to the non-stopwords, offering a means to standardize and normalize the data by reducing words to their base forms. These preprocessing steps serve as integral components of our methodology, carried out prior to the execution of computationally intensive models.

METHODOLOGY

The sentiment score was created by the integration of five distinct sentiment score extraction techniques: SentiWordNet, TextBlob, VADER, AFINN, and a calculation based on positive and negative word counts. These will be explained next.

SentiWordNet

SentiWordNet is a lexical database widely used for sentiment analysis tasks. It assigns sentiment scores to words based on their semantic orientations. SentiWordNet associates each word with three sentiment scores: positivity, negativity, and objectivity. These scores range from 0 to 1, representing the degree of sentiment expressed by a word. By leveraging SentiWordNet, sentiment analysis systems can determine the overall sentiment of a given text by calculating the sentiment scores of individual words and aggregating them. This valuable resource has proved helpful in various natural language processing applications, enabling researchers and developers to analyze sentiments and emotions in text data.

In this analysis, SentiWordNet is used to perform sentiment analysis on the 'text' feature. The code begins by defining a dictionary, 'pos_dict', which maps part-of-speech tags to WordNet's POS constants. The function 'token_stop_pos' takes a text input, tokenizes it, assigns part-of-speech tags to each word using the 'pos_tag' function, and creates a new list of word-pos pairs excluding stopwords. The next function, 'lemmatize', lemmatizes the words in the input text using the WordNet lemmatizer. If a part-of-speech tag is available for a word, it is passed to the lemmatizer to ensure accurate lemmatization. The lemmatized words are concatenated into a single string and returned. The 'sentiwordnetanalysis' function performs sentiment analysis on the lemmatized text. For each word-pos pair, it retrieves the corresponding synsets from WordNet and uses the SentiWordNet module to obtain the sentiment scores for the first synset. The positive score is subtracted from the negative score, and the result is accumulated in the 'sentiment' variable. The 'tokens_count' variable keeps track of the number of tokens processed. If the sentiment is positive, the function returns 1; if it's neutral, it returns 0; otherwise, it returns -1. Finally, the feature 'POS tagged' is processed using the 'sentiwordnetanalysis' function, and the sentiment analysis results are stored in a new column called 'senti_wordnet'.

VADER

Valence Aware Dictionary and sEntiment Reasoner (VADER) is a method present in the NLTK library. It is a lexicon and rule-based sentiment analysis tool which relies on a dictionary that maps lexical features to emotion intensities. These emotion intensities are then used to calculate sentiment scores. VADER is sensitive to both polarity and intensity and is well suited for generating sentiment scores for unlabeled data. It is also specifically attuned for social media text, which makes it an excellent tool to be used for Twitter data. VADER was applied to the Lemma feature, which was derived from preprocessing the text feature of each tweet. The output is a dictionary containing four key-value pairs: ‘neg’, ‘neu’, ‘pos’, and ‘compound’. The first three key-value pairs represent the proportion of the text which VADER determined to be negative, neutral, and positive, respectively, and ‘compound’ represents the overall sentiment of the text. Based on the ‘compound’ value, the text’s sentiment score was labeled as -1 for negative, 0 for neutral, and 1 for positive, and stored in the senti_vader column.

Afinn

Afinn lexicon contains a list of English words, where each word is given a rating from -5 to 5. A score from -5 to -1 indicates a negative sentiment, with -5 being a highly negative score; a score from 1 to 5 indicates a positive sentiment, with 5 being a highly positive score; and a score of 0 indicates a neutral sentiment. The word list itself contains 2,476 words, consisting of negatively scored, positively scored, and neutral scored English words. Afinn was applied to the Lemma column, similarly to VADER, and a score was generated for each tweet based on the collective sentiment score from the score of each identified word from Afinn lexicon that was present in the tweet. That score was then assigned and stored to the senti_afinn column. Also, each tweet was grouped categorically, being assigned “positive”, “negative”, or “neutral.”

Textblob

TextBlob is a powerful Python library that provides a simple and intuitive interface for performing sentiment analysis on text data. TextBlob makes this task easier by offering a pre-trained sentiment analyzer that can classify text sentiment with high accuracy. It uses a combination of machine learning techniques and linguistic rules to analyze the text and assign sentiment scores. TextBlob also provides additional features such as part-of-speech tagging and noun phrase extraction, making it a versatile tool for natural language processing tasks. Overall, TextBlob's sentiment analysis capabilities make it a valuable tool for understanding and extracting sentiment information from text data. TextBlob was used to analyze the sentiment of text data in the 'Lemma' feature.

Classification of tweets

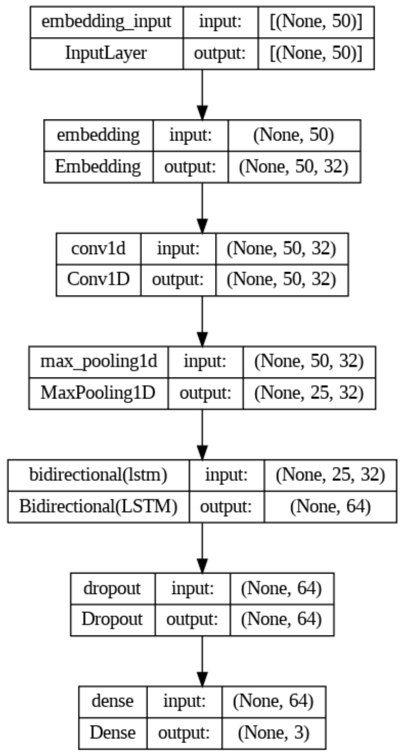
Following the acquisition of sentiment scores, we proceeded with text classification using the Long Short-Term Memory (LSTM) method. LSTM, an acronym for Long Short-Term Memory, is a specialized recurrent neural network (RNN) architecture designed to address the inherent limitations of conventional RNNs when dealing with extended sequential data. Its primary advantage lies in its capacity to effectively capture and retain information over long sequences.

In an LSTM, the network includes memory cells that can store and retrieve information over multiple time steps. These memory cells are equipped with specialized gates that control the flow of information, allowing the network to selectively update and forget information as needed. The three main gates in an LSTM are the input gate, forget gate, and output gate. The input gate determines how much new information is stored in the memory cells. The forget gate controls the extent to which existing information is discarded. The output gate regulates how much information from the memory cells is used to produce the output at each time step.

By using these gates and memory cells, LSTMs can effectively learn and model long-term dependencies in sequential data, making them suitable for tasks such as natural language processing, speech recognition, and time series prediction. The ability to retain relevant information over extended periods makes LSTMs particularly valuable when dealing with data that exhibits long-term dependencies or context.

Tensorflow Keras library was used to create the LSTM model. The model was constructed using a sequential architecture, encompassing several key layers. These layers include an embedding layer, a convolutional layer, a max pooling layer, a bidirectional LSTM layer, a dropout layer for regularization, and a dense output layer with softmax activation. The model was compiled utilizing the Stochastic Gradient Descent (SGD) optimizer and employed categorical cross-entropy loss as the loss function. For a visual representation of the model's architecture, refer to Figure 4.

Figure 4: Network architecture



RESULTS

Upon the implementation of the five sentiment models, we successfully derived the final sentiment feature in our dataset. This feature represents the mode among all the sentiment values generated. As depicted in Figure 5, it is evident that the models consistently yield comparable results. It is noteworthy that SentiWordNet classifies approximately 70% of the data as neutral, primarily occurring when the model is unable to confidently predict the sentiment of the text, leading to a neutral classification. However, when amalgamating all the values to form the final sentiment column, this prevalence of neutral classifications does not significantly impact our overall sentiment analysis, due to taking the mode of all sentiment scores for each tweet.

Figure 5: Sentiment analysis results

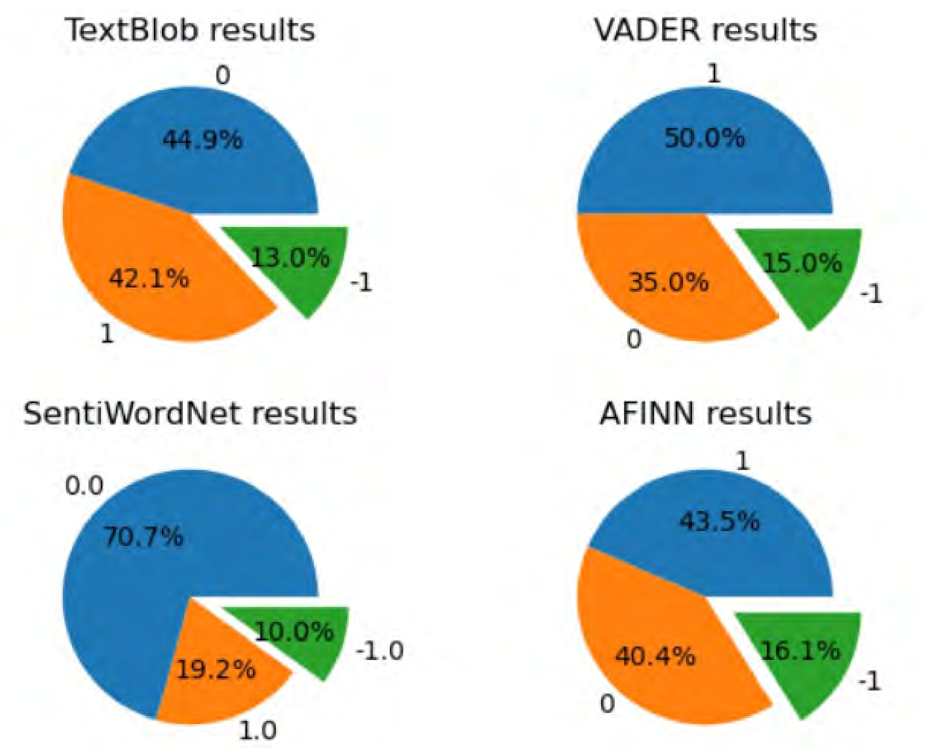
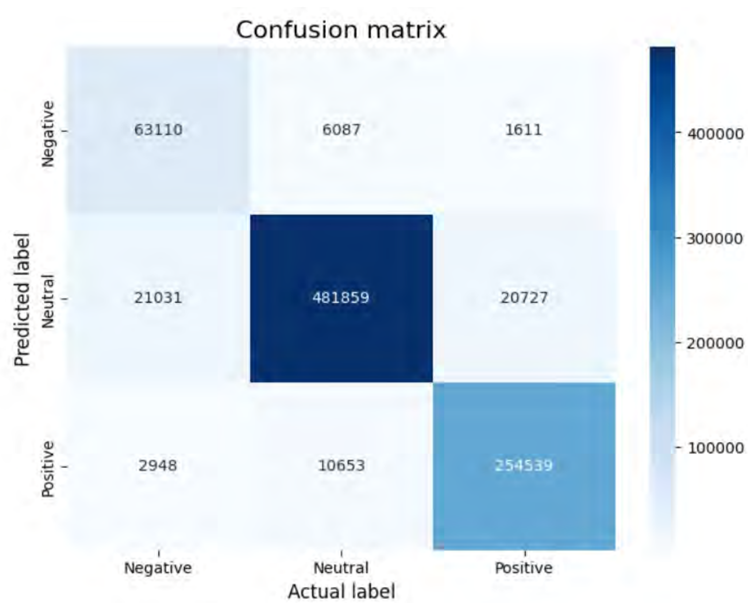


Figure 6: Confusion matrix for predicting the sentiment scores



.

LSTM performed overall very well with an accuracy of 0.9269 and an F1 score of 0.9270, which are very good scores. The confusion matrix can be seen in Figure 6.

CONCLUSION

This study embarked on a comprehensive analysis of cryptocurrency sentiment on Twitter, leveraging an amalgamation of sentiment scoring models and the powerful Long Short-Term Memory (LSTM) method for text classification. Our findings are promising, with LSTM exhibiting remarkable performance, boasting an accuracy of 0.9269, precision of 0.9317, recall of 0.9224, and an F1 score of 0.9270.

This research underscores the significance of harnessing Natural Language Processing techniques to glean insights from the continuous stream of Twitter data, shedding light on the intricate relationship between public sentiment and cryptocurrency price fluctuations. The utilization of multiple sentiment models and LSTM for text classification proved to be a robust approach, mitigating the limitations associated with relying solely on a single model. Moreover, the preprocessing steps involving stopwords removal and lemmatization played a crucial role in enhancing the quality of sentiment analysis.

REFERENCES

Baroiu, Alexandru-Costin, and Gabriela Dobrita Ene. “Twitter Sentiment and Bitcoin Price – Is There a Connection?” 2022 26th International Conference on System Theory, Control and Computing (ICSTCC), 2022, pp. 258–62, doi:10.1109/ICSTCC55426.2022.9931814.

Bansal, A., & Kumar, N. (2022). Aspect-Based Sentiment Analysis Using Attribute Extraction of Hospital Reviews. New Generation Computing, 40(4), 941–960.

Bello, A., Ng, S.-C., & Leung, M.-F. (2023). A BERT Framework to Sentiment Analysis of Tweets. *Sensors (14248220)*, *23*(1), 506.

(Cheng) Long, S., Lucey, B., Xie, Y., & Yarovaya, L. (2023). “I just like the stock”: The role of Reddit sentiment in the GameStop share rally. *Financial Review*, 58, 19– 37.

Chong, W. Y., Selvaretnam, B., & Soon, L.-K. (2014). Natural language processing for sentiment analysis: An exploratory analysis on Tweets. *2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology*. https://doi.org/10.1109/icaiet.2014.43

Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on Deep Learning: A Comparative Study. *Electronics*, *9*(3), 483. https://doi.org/10.3390/electronics9030483

Diana, T. (2023). Leveraging sentiment analysis as a predictor of risk in community engagement. Journal of Airport Management, 17(2), 182–196.

Edgari, Eric, et al. “The Impact of Twitter Sentiment Analysis on Bitcoin Price during COVID-19 with XGBoost.” 2022 5th International Conference on Computing and Informatics (ICCI), 2022, pp. 337–42, doi:10.1109/ICCI54321.2022.9756123.

Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users' Comments. Sensors (14248220), 23(4), 1748.

Haritha G. B., and Sahana N. B. "Cryptocurrency Price Prediction Using Twitter Sentiment Analysis." Natural Language Processing, Information Retrieval and AI , Academy, 2023, doi:10.5121/csit.2023.130302.

Hasan, Md. R., Maliha, M., & Arifuzzaman, M. (2019). Sentiment analysis with NLP on Twitter data. *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*. https://doi.org/10.1109/ic4me247184.2019.9036670

Khurshid, Abdul Rehman. Cryptocurrency Price Prediction Using Sentiment Analysis. 2021.

Kim, M., Park, E. L., & Cho, S. (2018). Stock price prediction through sentiment analysis of corporate disclosures using distributed representation. Intelligent Data Analysis, 22(6), 1395–1413.

Mukherjee, S., & Bala, P. K. (2017). Detecting sarcasm in customer tweets: An NLP based approach. *Industrial Management & Data Systems*, *117*(6), 1109–1126. https://doi.org/10.1108/imds-06-2016-0207

Pota, M., Ventura, M., Fujita, H., & Esposito, M. (2021). Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets. *Expert Systems with Applications*, *181*, N.PAG.

Purohit, A., & Patheja, P. S. (2023). Product review opinion based on sentiment analysis. Journal of Intelligent & Fuzzy Systems, 44(2), 3153–3169.

Shanmugavadivel, K., Sathishkumar, V. E., Raja, S., Lingaiah, T. B., Neelakandan, S., & Subramanian, M. (2022). Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data. *Scientific Reports*, *12*(1), 1–12.

Shukla, D., Chandra, G., Pandey, B., & Dwivedi, S. K. (2022). A comprehensive survey on sentiment analysis: Challenges and future insights. *Journal of Intelligent & Fuzzy Systems*, *43*(6), 7733–7763.

Tung, Pham Thai Hoang, et al. SemiMemes: A Semi-Supervised Learning Approach for Multimodal Memes Analysis. 2023.

Weerasooriya, T., Perera, N., & Liyanage, S. R. (2016). A method to extract essential keywords from a tweet using NLP Tools. *2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer)*.

Zhang, X., Zhou, H., Yu, K., Wu, X., & Yazidi, A. (2023). Tsetlin Machine for Sentiment Analysis and Spam Review Detection in Chinese. Algorithms, 16(2), 93.

DECISION SCIENCES INSTITUTE

Board Gender Diversity and Workplace Safety: Evidence from Quasi-Natural Experiments

Md Ismail Haidar
University of Texas Rio Grande Valley
E-mail: mdismail.haidar01@utrgv.edu

Saha Iqbal Hossain
Texas A&M International University
E-mail: sahahossain@dusty.tamiu.edu

ABSTRACT

This paper examines the impact of board gender diversity on workplace safety in an international setting from 2003 to 2019. We exploit the introduction of gender quotas on corporate boards as a quasi-natural experiment to examine the causal effect of board gender diversity on workplace safety. Using a sample of 13,124 firm-year observations from 48 countries, the study finds that injury rates increase with gender quotas on the board. Cross-sectionally, the effect is stronger for financially constrained firms and for those with high workloads and low workplace investment. However, the effect is weaker for firms in countries with the good institutional quality and high union representation. The results remain robust to alternative regression specifications and alternative series of workplace injuries and board gender diversity.

KEYWORDS: Workplace Safety, Gender Quotas, Board Gender Diversity, Difference-in-Differences

DECISION SCIENCES INSTITUTE

Building robust supply chains through digitalization. The role of IT-enabled dynamic capabilities

Caleb Amankwaa Kumi
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
caleb.kumi3@gmail.com / ca.kumi@knust.edu.gh

Francis Kofi Andoh-Baidoo
University of Texas Rio Grande Valley, Edinburg, TX, USA
francis.andohbaidoo@utrgv.edu

David Asamoah
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
dasamoah.ksb@knust.edu.gh

Emmanuel Kweku Quansah
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
equansah@hotmail.com

John Serbe Marfo
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
serbemarfo@gmail.com

ABSTRACT

This study investigates the mechanism through which supply chain (SC) digitalization enhances supply chain robustness. Drawing from the resource orchestration theory and the IT-enabled dynamic capabilities view, we develop a theoretical model that is tested with data obtained from 286 Ghanaian firms. Results from the study show that SC digitalization has a direct positive relationship with organizational sensing and learning dynamic capabilities. Further, the results indicate that IT-enabled dynamic capabilities (sensing, learning, and reconfiguration) mediate the relationship between SC digitalization and supply chain robustness. We offer theoretical and practical implications for the extant literature and provide further avenues for research.

KEYWORDS: Disruption, Supply chain digitalization, Supply chain robustness, Dynamic capabilities, Resource orchestration

INTRODUCTION

Contemporary supply chains are complex systems due to the presence of a highly dynamic business environment, multi-layered supply chain actors, globalization, and disruption risks (Qrunfleh et al., 2023). The complexities in the supply chain make it vulnerable and susceptible to disruptions. In recent times, supply chains worldwide have experienced an unprecedented series of shocks caused by the COVID-19 virus outbreak and global pandemic, a new instigator of supply chain disruptions (Chesbrough, 2020; Sarkis et al., 2020). Unfortunately, the impact of

COVID-19 may not stop soon (Pimenta et al., 2022). A fundamental question is what businesses can do to survive in today's complex business environment rife with disruptions (Xie et al., 2022). Thus, it has been on the agenda of many organizations to devise some capabilities to survive these supply chain disruptions and uncertainties. A notable disruption mitigation capability that has been touted by the extant literature is supply chain robustness which is identified as a proactive approach to dealing with supply chain disruptions (Brandon-Jones et al., 2014; El Baz and Ruel, 2021; Dolgui, 2020). Supply chain robustness refers to the ability of an organization to proceed with its planned operation despite the occurrence of a disruption (Simchi-Levi et al., 2018). While robustness has been recognized as a crucial topic, the subject has not been studied extensively (El Baz and Ruel 2021; Alvarenga et al., 2023).

Arguably, supply chain digitalization is a key resource that can be orchestrated by organizations to develop capabilities for surviving disruptions (Yang et al., 2021; Alvarenga et al., 2023). Supply chain digitalization is defined as the socio-technical process of using emerging digital technologies to transform and improve supply chain activities at the focal firm and across the upstream and downstream supply chain (Ehie and Ferreira, 2019; Hennelly et al., 2020; Saarikko, 2020). Digitalization facilitates the boundary-spanning activities of supply chains by fostering inter and intra-organizational information sharing (Lui et al., 2016). Thus, digitalization empowers supply chains with real-time access to relevant information (Richey and Davis-Sramek, 2021) to manage complexities and disruptions. Further, it enhances the information processing and sensing capabilities of firms to deal with unforeseen events (Gu et al., 2021).

The experiences from the recent global pandemic suggest the need for firms to pursue supply chain robustness (Alvarenga et al., 2023). This has instigated a paradigm shift towards digitalization and robustness as disruption mitigation mechanisms. While supply chain robustness has been recognized as a crucial topic, the subject has not been studied extensively (El Baz and Ruel, 2021; Alvarenga et al., 2023). Very little is known on how supply chain robustness can be nurtured by organizations. Discussions on the impact of supply chain digitalization on robustness remain anecdotal with very limited empirical evidence on how supply chain digitalization may be leveraged to enable supply chain robustness. Given the potentially significant role of supply chain digitalization and robustness in improving an organization’s ability to withstand disruptions, this paucity of empirical evidence needs to be addressed.

While research generally supports the proposition that supply chain digitalization enhances disruption mitigation capabilities (Cui et al., 2022), some scholars have argued that the deployment of digital technologies does not always yield the anticipated outcomes (Ye et al., 2022). This infers that the implementation of supply chain digitalization will not automatically guarantee supply chain robustness. Hence, the mechanisms through which supply chain digitalization affects supply chain robustness is an important research gap that must be addressed. Supply chain digitalization is essential but has limited value if the intervening factors that facilitate its performance outcome are not considered (Cui et al., 2022).

Consequently, this study seeks to answer two closely related research questions: (1) Does supply chain digitalization result in supply chain robustness? and (2) Through what mechanism do IT-enabled capabilities mediate the relationship between digitalization and supply chain robustness? Accordingly, this study seeks to address the lacuna above by drawing from the resource orchestration theory and the IT-enabled dynamic capabilities view to explain the mechanism through which supply chain digitalization enhances supply chain robustness. The study explains how supply chain IT-enabled dynamic capability acts as a facilitating mechanism in the relationship between supply chain digitalization and supply chain robustness. IT-enabled dynamic capabilities are defined as a firm's abilities to leverage its IT resources and IT competencies, in combination with other organizational resources and capabilities, to address rapidly changing business environments" (Mikalef et al., 2021).

This research yields several theoretical contributions while offering noteworthy managerial implications. First, it contributes to the digitalization and supply chain robustness literature by responding to a call for more empirical research on the topic. Specifically, the study empirically demonstrates the significant role supply chain digitalization plays in enabling firms to build shock absorbers to mitigate the impact of supply chain disruption. Second, this study contributes to theory by explaining the mediating mechanisms through which supply chain IT-enabled dynamic capabilities transfer the effect of supply chain digitalization on supply chain robustness.

THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

Resource orchestration theory

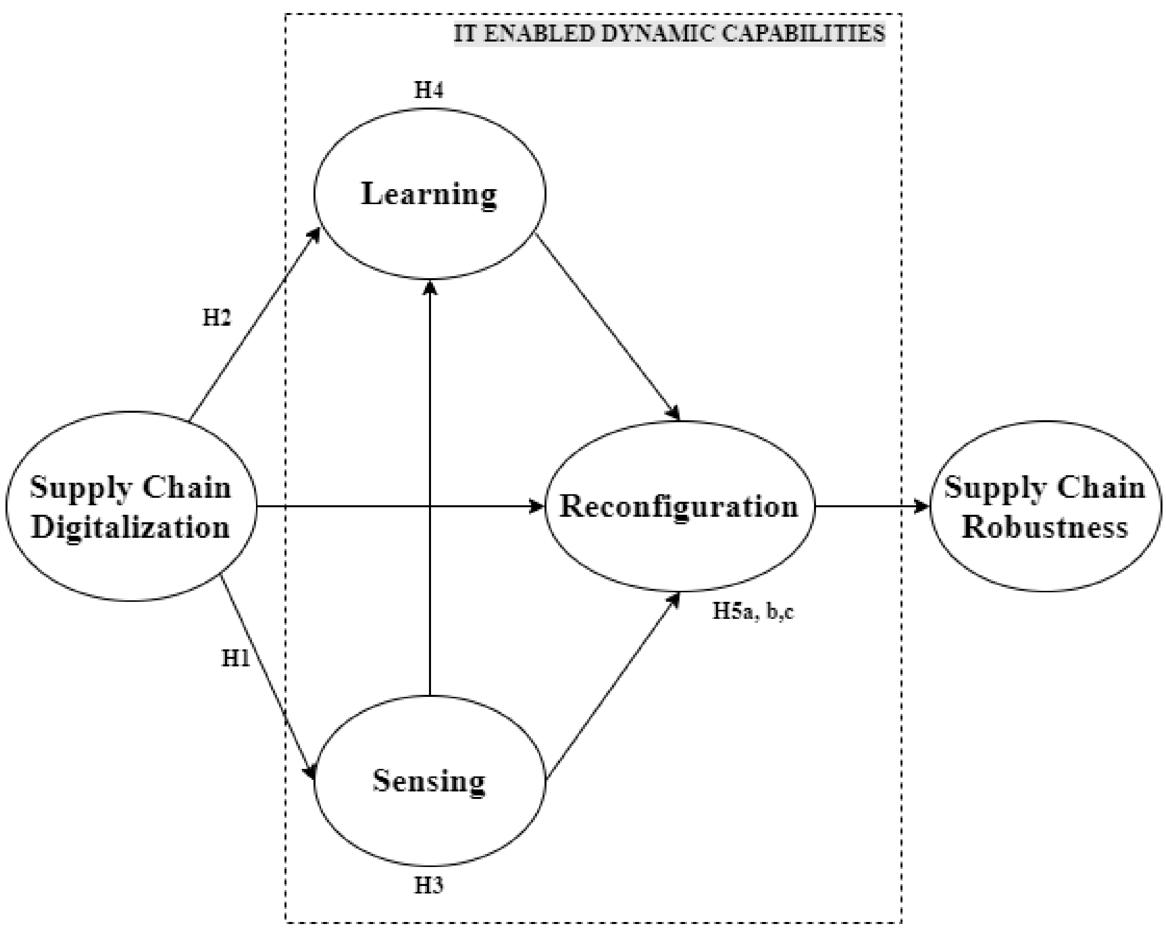
Resource orchestration encompasses the combined efforts of resources, capabilities, and managerial competence to facilitate the deployment of resources (Sirmon et al., 2007). Resource orchestration involves not only the integration of mobilized resources but also the management of these resources (Chadwick et al., 2015). The resource orchestration theory can be viewed as an extension of the resource-based view (RBV) theory, as it lays the foundation for effectively structuring, bundling, and leveraging firm resources to enhance resource accumulation (Wong et al., 2015). The resource orchestration theory emphasizes the extent to which a firm can fully utilize its available resources (Asiaei et al., 2021). The theory underscores the importance of resource mobilization, as mobilized resources can be integrated into functional structures to improve synchronization, alignment, and utilization (Helfat et al., 2007). Therefore, it is crucial for firms to fully exploit the potential of their resources and capabilities by deploying them in a complementary manner (Burin et al., 2020; Liu et al., 2016). From the theoretical perspective of the resource orchestration theory (ROT), supply chain digitalization is considered a significant organizational resource that mobilizes and integrates the activities of interconnected firms within a supply chain (Asamoah et al., 2021; Yang et al., 2021). We posit that supply chain digitalization constitutes a vital organizational resource that can be effectively orchestrated to cultivate IT-enabled dynamic capabilities essential for bolstering supply chain robustness.

IT-enabled dynamic capability

IT-enabled dynamic capabilities refer to a firm's ability to leverage IT resources and competencies in combination with other organizational resources to navigate changing business environments (Mikalef and Pateli, 2017). In today's dynamic business world, firms face sustainability-related challenges in global logistics operations, and emerging digital technologies are crucial in addressing these challenges (Li et al., 2022; Mikalef et al., 2021). Information Technology (IT) plays a vital role in both operational and strategic aspects of firms' value-creation processes (Bag et al., 2020; Felipe et al., 2019). However, simply investing in IT does not guarantee a sustainable competitive advantage (Ye et al., 2022; Gupta and George, 2016).

Firms must develop organizational capabilities through the effective utilization of IT to demonstrate superior performance (Brinch et al., 2021; Mikalef et al., 2021; Bag et al., 2020). The IT-enabled dynamic capability view argues that IT is valuable only when effectively harnessed to enable or strengthen an organizational capability. Despite studies exploring IT-enabled capabilities, limited empirical research exists on the enabling effect of IT on dynamic capabilities and its impact on firm performance outcomes (Majhi et al., 2022). This study aims to contribute by explaining how IT-enabled dynamic capabilities foster the relationship between supply chain digitalization and supply chain robustness. Three key capabilities will be examined: sensing, seizing, and reconfiguration. Sensing capability involves identifying and interpreting opportunities and threats in the environment (Teece, 2007). Learning capability entails acquiring and exploiting new knowledge for informed decision-making (Mikalef and Pateli, 2017). Reconfiguration capability entails a firm's ability to adapt strategically and demonstrate agility in changing operational modes (Teece, 2007; Mikalef and Pateli, 2017).

Figure 1: Research model



Supply chain digitalization and sensing capability

Sensing, which refers to the ability to acquire innovative ideas from external and diverse sources to produce essential innovation (Camison, 2012), plays a crucial role in the dynamic supply chain environment. In this context, sharing information through integrated systems within and among firms is essential (Deepu & Ravi, 2021). Accurate and real-time information sharing among supply chain partners contributes to effective supply chain management (Asamoah et al., 2021; Kumi et al., 2021) and enables firms to sense their business environment, collect data, and exploit it to enhance firm growth (Arunachalam et al., 2018; Agyei-Owusu et al., 2021). Sensing goes beyond rapid adaptation to external sources; it also involves the internal exploitation of existing knowledge and individuals' past experiences (Ma et al., 2015). Through supply chain digitalization, firms can augment their internal resources and capabilities with external resources available to their partners (Asamoah et al., 2021). This enables them to identify threats and opportunities within the working environment. In a supply chain digitalization system, firms engage in communication, integration, and intelligence building, which enhances their ability to identify unique sources of supply chain disruptions. Viewing it from a resource orchestration perspective, supply chain digitalization empowers firms to identify threats and opportunities in the external environment through the integration and management of mobilized resources (Chadwick et al., 2015). This allows firms to constantly search for opportunities that align with their available resources and capabilities (Teece, 2007). In summary, integrating supply chain digitalization into firms' practices enhances their ability to sense potential threats and opportunities from various heterogeneous contexts. By deploying effective strategies to mitigate threats and capitalize on opportunities, firms are more likely to leverage supply chain digitalization for improved sensing capabilities. Therefore, we hypothesize the following:

H1: Supply chain digitalization has a positive relationship with firms' sensing capabilities.

Supply chain digitalization and learning capability

Learning is a fundamental higher-order capability (Teece, 2007) that empowers firms to acquire and utilize sufficient knowledge for creating and transforming their capability and resource base (Zahra & George, 2002). The adoption of supply chain digitalization, which incorporates communication, integration, and intelligence-building, enhances firms' competence (Sirmon et al., 2007) within the dynamic environment. This allows firms to study and understand the threats and opportunities they face. Given the complexities inherent in supply chains, it is crucial for firms to gain insights from their environment. This can be achieved through inter-organizational information sharing, integration, and intelligence-building. By sharing predetermined mechanisms among supply chain partners, firms can analyze the potential impact of these mechanisms on supply chain disruptions and extract valuable insights on how to minimize their effects or even transform them into new opportunities. From a resource orchestration perspective, mobilized resources that are integrated into functional structures should be synchronized and aligned based on specific directions of use. This directional insight can be facilitated through the strategic learning process (Helfat et al., 2007). Supply chain digitalization plays a critical role in mobilizing resources from different sources and, when effectively deployed, enhances firms' dynamic capabilities through strategic learning. By integrating supply chain digitalization into their practices, firms are better positioned to acquire and utilize sufficient knowledge, enabling them to create and reform their capabilities and resources to align with identified threats and opportunities. Accordingly, the study hypothesizes that:

H2: Supply chain digitalization is positively related to firms’ strategic learning capabilities.

The mediating role of sensing capability and strategic learning between supply chain digitalization and reconfiguration

In a dynamic business environment, a firm’s ability to sense and filter strategic opportunities concerning its business model is an important step in addressing the needs of the changing environment (Teece, 2012). In building a firm’s supply chain digitalization capacities, it is essential to authenticate its business model, monitor competitors’ models, and scan for external and internal threats to reconfigure business processes (Gelhard et al., 2016). Through sensing enabled by supply chain digitalization, firms can generate new information as well as monitor market opportunities. In a highly competitive business environment, firms with high awareness of their business models as well as competitors’ business models are better positioned to discover new business models which could help them to capitalize on new opportunities and strategically reconfigure their business than companies with low awareness (Pavlou & El Sawy, 2011). Strategic learning remains essential to the reconfiguration and innovation capabilities in organizations (Van den Bosch et al., 1999). According to (Muneeb et al., 2019), the ability to enrich and widen firms’ insight into available opportunities and reconfigure avenues is a strategic move to identify new reconfiguring opportunities. Learning has been considered central to the basis of developing dynamic capabilities (Ambrosini et al., 2009; Teece, 2007). Learning as a process involves repetition and experimentation which helps to improve earlier processes by identifying more efficient processes. It is argued that resource reconfiguration seems nearly unrealistic when functional learning mechanisms that improve upon earlier mistakes and reinvent knowledge structures based on learned experiences are missing (Galunic and Rodan, 1998). Strategic learning enhances firms’ innovativeness and enhances them to adapt to the changing needs of the environment such as technological changes and market fluctuations as well as facilitating the modification and transformation of business models (Gelhard et al., 2016). From the resource orchestration perspective, resource combination may be repetitively reconfigured to identify useful assets and efficient business processes necessary for competitive advantage (Muneeb et al., 2019). Moreover, organizations that engage in strategic learning are less likely to experience high levels of organizational sluggishness (Levinthal, 1991), which helps them to achieve reconfiguration. Accordingly, strategic learning enabled by supply chain digitalization fosters both the effective selection and the actual development of business models toward competitive advantages (Teece, 2007). Some studies (Asamoah et al., 2021; Mandal & Dubey, 2021) have reported on the important role of digitalization in supply chains. Amid supply chain disruptions, these important roles may not easily be seen, especially in a dynamic environment where systems are not strong enough to weaken the various threats. Hence, it is argued that the complexities in supply chains might lessen the effectiveness of supply chain digitalization on reconfiguration capabilities. In view of this, sensing and learning become important capabilities to consider to help minimize the potency of various threats in supply chains towards reconfiguration. Moreover, the direct link between supply chain digitalization and reconfiguration seems unrealistic, hence the need to channel the effect of supply chain digitalization through organizational sensing capability and strategic learning capability, which have the potency to minimize various threats and to capitalize on the opportunities becomes an important link to consider. Accordingly, it is hypothesized that:

H3: Sensing capability significantly mediates between supply chain digitalization and reconfiguration capability.

H4: Strategic learning capability significantly mediates between supply chain digitalization and reconfiguration capability.

The mediating role of reconfiguration dynamic capabilities in the relationship between supply chain digitalization and robustness in supply chain

The combined effort of the various dynamic capabilities, which are sensing, strategic learning, and reconfiguration are assumed to channel the effect of supply chain digitalization on supply chain robustness. From the resource orchestration theory, supply chain digitalization serves as an important organizational resource (Asamoah et al., 2021) and this gives a new role to management in expanding, building, etc. the organizational resources in such a way as to optimize performance (Sirmon et al., 2011). In such a dynamic environment, building and expanding resources may not easily be attained without developing efficient dynamic capabilities to facilitate it (Badewi et al., 2018). To achieve the full potential of these capabilities, a synergistic approach is required to synchronize the various dynamic capabilities (sensing, strategic learning, and reconfiguration) and use them complementary (Rojo et al., 2020). From the lens of orchestration theory, developing extra resources will lead to a capability that leads to higher performance outcomes (Davis-Sramek et al., 2015). Thus, through an orchestration mechanism, performance outcomes can be transformed from one level to the next by purposely spreading firms' capabilities (Badewi et al., 2018). Considering the nature of the competition and the dynamic environment, building a robust supply chain may seem difficult to come by even with the use of supply chain digitalization. Robustness in the supply chain is considered a proactive strategy to cope with changes, turbulences, or disruptions (El Baz & Ruel, 2021; Chowdhury and Quaddus, 2017). Thus, robustness may not be too difficult to achieve if one is fully aware of the changes to come or the disruptions. Hence awareness of the business model is one essential strategy to help strengthen supply chain robustness and this awareness can be made possible through sensing, a dynamic capability. Strategic learning as another lens of dynamic capability helps firms to innovate and adapt to a changing environment, whereas reconfiguration enables firms to adapt more quickly and effectively, creating a stream of temporary competitive advantages (Gelhard et al., 2016). In building robust supply chains, the direct link from supply chain digitalization to supply chain robustness may seem unrealistic. Accordingly, building reconfiguration dynamic capabilities through supply chain digitalization, sensing, and strategic learning is argued to be an effective means through which supply chains can cope with changes, turbulences, or disruptions through robustness. Hence it is hypothesized that:

H5a-c: Reconfiguration dynamic capabilities significantly mediate between supply chain digitalization, sensing, strategic learning, and supply chain robustness.

METHODOLOGY

Research design and setting

We conducted a cross-sectional survey to examine our hypotheses, selecting hospitals in Ghana as our sample. Ghana, being a developing country, poses challenges in acquiring objective secondary data for the variables of interest (Kull et al., 2018; Essuman et al., 2023). While cross-sectional data limits our ability to establish causality, Rindfleisch et al. (2008) argue that it can still be utilized to test explanatory models based on relevant theories. Thus, our decision to use cross-sectional data is justified. By selecting Ghana as our research setting, we capture the unique characteristics of an emerging economy that has faced significant disruptions, including the COVID-19 pandemic, the Russian-Ukraine war, power cuts, and natural disasters such as floods and fire outbreaks. Also, Ghana is experiencing considerable digitalization in the health sector. Thus, data obtained from this context enables us to gain a nuanced understanding of the subject matter of supply chain robustness.

Measure development

The study employed a questionnaire-based survey method due to its advantages in facilitating the generalizability of outcomes, ease of replication, and the ability to simultaneously investigate a large number of factors (Pinsonneault & Kraemer, 1993). The items used to measure the constructs in our research model were adopted from prior studies. The constructs and corresponding items are operationalized on a 7-point Likert scale. The Likert scale for digitalization was anchored from 1 = “extremely low” to 7 = “extremely high”. Further the Likert scale for IT-enabled dynamic capabilities (sensing, learning, and reconfiguring) ranged from 1 = “not effective at all” to 7 =” highly effective”. Finally, the scales for supply chain robustness ranged from 1 = “strongly disagree” to 7 = “strongly agree”. The items for measuring IT-enabled dynamic capabilities (sensing, learning, and reconfiguring) were adopted from Mikalef et al., 2021). Sources for measuring the remaining items are as follows: Digitalization (Yang et al., 2021; Xu et al., 2013) and supply chain robustness (Brandon-Jones 2014; El Baz and Ruel, 2021). To mitigate alternative explanations, we controlled for firm size, firm age, and firm revenue. Before administering our questionnaire, we sought the expertise of three scholars specializing in supply chain and information systems. They carefully reviewed the operational definitions and indicators of the constructs we intended to measure. Based on their valuable feedback, we eliminated indicators that exhibited poor face validity and revised items that were found to be ambiguous. We conducted a pretest by administering it to 30 senior executives in the health (Eg. operations managers, IT managers, and supply chain managers) who were participating in an Executive MBA program. After analyzing the pilot test data, no major concerns arose regarding the questionnaire. Table 1 shows the final list of indicators and their psychometric properties.

Data Collection

The target firms for the survey were health facilities that operate in Ghana. Health facilities offer an ideal context for this study. First, the complexity and dynamics of healthcare supply chains, combined with the criticality of healthcare products, make them a practical and relevant setting for research. Moreover, the regulatory compliance requirements and collaborative nature of the healthcare industry further enhance the suitability of health facilities as a context for examining the impact of supply chain digitalization on robustness. The study considered only firms that have been in business for more than 3 years and have adopted digital supply chains. Based on a list obtained from the Ministry of Health, 400 health facilities were randomly selected and contacted. Out of the firms approached 320 were willing to partake in the study. Accordingly, we administered 320 questionnaires using a team of trained field workers working directly under our supervision. Consistent with past digitalization and supply chain robustness research (Alvarenga et al., 2023), we targeted only senior-level managers (Eg. operations manager, supply chain manager, and IT managers, among others). We obtained a total of 296 returned questionnaires, of which 10 were discarded as they were incomplete, leaving 286 valid responses. Power analysis was used to test the adequacy of our sample size. Using a recommended power of 0.8, a probability of error of 0.05, and a medium effect size of 0.3 (Cohen, 2013), the required sample size for the study was 102. Therefore the 286 responses used for the study were sufficient. In estimating the effect of common method bias, Harman's one-factor test was employed. The results showed that 35.2% of the variance in the model was accounted for by the single factor extraction which is significantly less than the 50% threshold, thereby indicating that common method bias was not a serious problem in the study. Non-response bias was assessed using the method recommended by Armstrong and Overton (1977), which compared early and late respondents on firm size and industry. The t-test results indicate no significant statistical difference ($p < 0.05$) between the category means for firm revenue and employee size, indicating that non-response bias may not be a concern in this study.

Table 1: Measurement items, item loadings, and psychometric properties of constructs

Items	Loadings	CA	CR	AVE
Supply Chain Digitalization Kindly indicate the degree to which your organization transacts with its suppliers/customers using digital supply chain systems.				
The proportion of suppliers that the firm transacts with through the digital supply chain system	0.88	0.948	0.958	0.793
The proportion of transactions (volume) that the firm conducts with its suppliers through the digital supply chain system	0.915			
The proportion of transactions (number of activities) that the firm conducts with its suppliers through the digital supply system	0.899			
The proportion of customers that the firm transacts with through the digital supply chain system	0.87			
The proportion of transactions (volume) that the firm conducts with its customers through the digital supply chain system	0.899			
The proportion of transactions (number of activities) that the firm conducts with its customers through the digital supply chain system	0.88			
Sensing Please indicate how effective your company is in using IT systems for the following purposes				
Scanning the environment and identifying new business opportunities	0.904	0.92	0.942	0.803
Reviewing our product/service development efforts to ensure they are in line with what the customers want	0.877			
Implementing ideas for new products/services and improving existing products/services	0.922			
Anticipating discontinuities arising in our business domain by developing greater reactive and proactive strength	0.881			
Learning Please indicate how effective your company is in using IT systems for the following purposes.				
Identify, evaluate, and import new information and	0.866	0.926	0.948	0.819

knowledge				
Transform existing information into new knowledge	0.938			
Assimilate new information and knowledge	0.918			
Use accumulated information and knowledge to assist decision making	0.896			
Reconfiguring Please indicate how effective your company is in using IT systems for the following purposes.				
Adjusting our business processes in response to shifts in our business priorities	0.879	0.904	0.933	0.777
Adjusting for and responding to unexpected changes easily	0.913			
Easily adding an eligible new partner that you want to do business with, or removing ones which you have terminated your partnership.	0.829			
Reconfiguring our business processes in order to come up with new productive assets	0.904			
Supply Chain Robustness				
Our supply chain can remain effective and sustained even when disruptive events occur (e.g. natural disasters, labour strikes, fire, industrial accidents, shortages in the supply markets)	0.84	0.858	0.898	0.639
Our supply chain performs well over a wide variety of possible scenarios	0.841			
Our supply chain can avoid or minimize risk occurrence by anticipating and preparing for them	0.801			
When changes occur, our supply chain grants us sufficient time to consider a reasonable reaction	0.723			
Our supply chain can absorb a significant level of negative impacts from recurrent risks	0.786			

RESULTS

The demographic data revealed that cumulatively, 43.9 % of the organizations which took part in this study had employed more than 100 employees, falling under the classification of large firms. Also, 27.5% had employed between 30 to 59 employees with a few 10% employing less than 6 employees. With regards to how long the firms have been in operation, the majority of the firms representing 51.2% had been in operation for more than 5 years. Concerning the estimated annual revenues, the majority of the respondents resenting 50.4% of responding firms had revenue levels of more than $10,000.

Measurement model results

The partial least squares structural equation modeling (PLS-SEM) approach was adopted to examine the research model of the study. Following the recommendations of Hair et al. (2019) the measurement model was analyzed by assessing the convergent validity and discriminant validity of the model.

First, convergent validity is examined by measuring the composite reliability of constructs, average variance extracted (AVE), and factor analysis. The results show that the item loadings were greater than 0.708 giving indications of good items reliability. Again, the Composite reliability values exceeded the suggested 0.7 threshold and the AVEs of all the constructs are higher than 0.5 as required, altogether indicating that the model has acceptable convergent validity. These results are summarized in Table 1.

Having confirmed convergent validity, the study further tested for discriminant validity. In establishing discriminant validity, we first used the Fornell-Larcker criterion. The Fornell-Larcker criterion is assessed by comparing the square root of the AVE of each construct with the correlation between constructs and to confirm if the former is larger than the latter (Hair et al. 2019). In Table 2 below, the diagonal values in bold which represent the square root of the AVE of constructs are larger than the off-diagonal values which are correlations among constructs, establishing discriminant validity.

Further, an HTMT test was undertaken as an additional test of discriminant validity. The HTMT test approach indicates that HTMT values must be significantly less than 1, with a value of less than 0.85 ideal (Hair et al., 2019). Table 3 indicates that the highest HTMT value is 0.88, confirming the model has sufficient discriminant validity.

Table 2: Fornell-Larcker Criterion

Constructs	1.	2.	3.	4.	5.
1. Supply Chain Digitalization	**0.891**				
2. Learning	0.55	**0.905**			
3. Robustness	0.362	0.3	**0.799**		
4. Sensing	0.25	0.255	0.668	**0.896**	
5. Reconfiguration	0.561	0.783	0.396	0.276	**0.882**

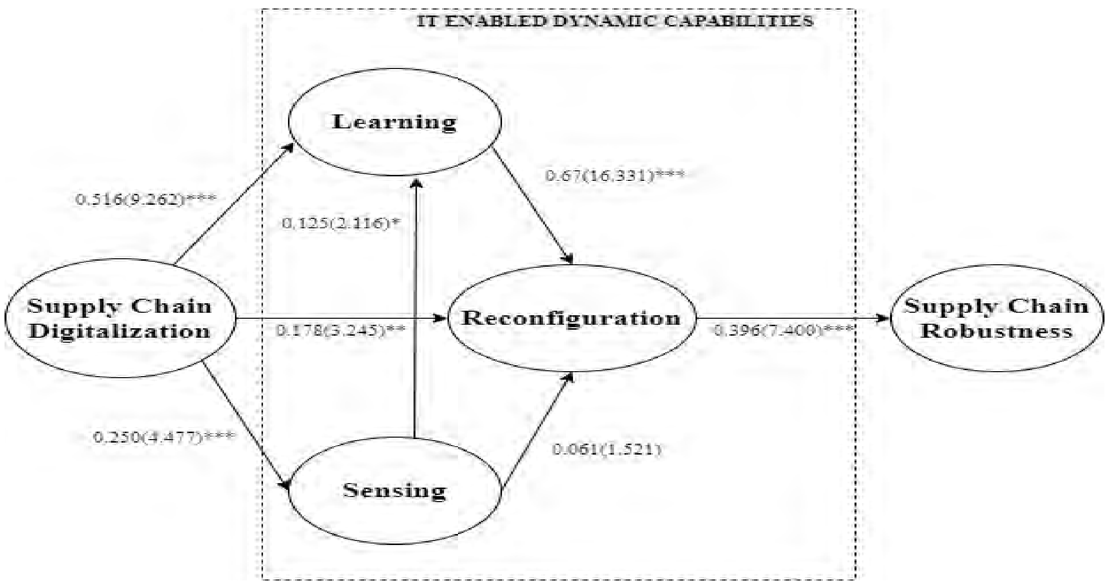
Table 3: HTMT

Constructs	1.	2.	3.	4.
1. Digitalization				
2. Learning	0.585			
3. Robustness	0.401	0.334		
4. Sensing	0.256	0.263	0.755	
5. Reconfiguration	0.605	0.855	0.448	0.288

Structural model results

Having confirmed the soundness of the measurement model, the structural model and hypothesized relationships were examined. First, the research model's in-sample explanatory power was examined by assessing the r-squared (R^2) of the endogenous variables, which ranged from 0.317 for sensing, 0.062 for learning, 0.640 for reconfiguration, and 0.157 for supply chain robustness. Also, examining the predictive relevance of the model revealed Q^2 values ranging from 0.045 to 0.491, confirming the predictive relevance of the model (Hair et al., 2019). To examine the hypotheses for the direct paths, the path coefficient and p-values were assessed. Additionally, the mediated paths were examined by following the guidelines of Nitzl et al. (2016). Consistent with previous information systems and supply chain literature, we examined the impact of control variables (firm size, firm age, and firm revenue) on the dependent variable (supply chain robustness): however, their relationship to the dependent variable was not significant. The results of the hypotheses tests are summarized in Table 4 below.

Figure 1: Model Results



Note: ***p<0.001; **p<0.01; *p<0.1

Table 4: Hypotheses Results

Paths	Path Coefficient	T Statistics	P Values	Decision
H1: SCD -> Sense	0.250	4.477	0.000	Supported
H2: SCD -> Learn	0.519	9.262	0.000	Supported
H3: SCD -> Sense -> Reconfig	0.015	1.391	0.165	Not Supported
H3: SCD -> learn -> Reconfig	0.347	7.583	0.000	Supported
H5a: SCD -> Reconfig ->Robust	0.070	2.981	0.003	Supported
H5b: Sense -> Reconfig ->Robust	0.024	1.338	0.181	Not Supported
H5c: Learn -> Reconfig ->Robust	0.265	7.009	0.000	Supported

Note: *SCD=Supply Chain Digitalization; Sense=Sensing Capability; Learn= Learning Capability; Reconfig = Reconfiguration Capability; Robust = Supply Chain Robustness*

As depicted in Table 4 and Fig. 2, five out of our seven hypothesized paths were empirically supported. The results show that supply chain digitalization is positively related to organization sensing capabilities ($\beta = 0.250$, $t = 4.477$, $p < 0.05$) and learning capabilities ($\beta = 0.519$, $t = 9.262$, $p < 0.05$). We tested for mediation using the bootstrapping approach in Smart PLS (Hair et al., 2016). We found that learning capability mediates the relationship between digitalization and robustness ($\beta = 0.347$, $t = 7.583$, $p < 0.05$). Contrary to our postulation, the results showed that sensing does not mediate the relationship between digitalization and reconfiguration ($\beta = 0.015$, $t = 1.391$, $p > 0.05$). The results from the mediating analysis revealed that reconfiguration mediates the relationship between supply chain digitalization and supply chain robustness ($\beta = 0.070$, $t = 2.127$, $p < 0.05$). Further, the results show that reconfiguration mediates the relationship between learning and supply chain robustness ($\beta = 0.265$, $t = 7.009$, $p < 0.05$). Contrary to our hypotheses, the results show that reconfiguration does not mediate the relationship between sensing capabilities and supply chain robustness ($\beta = 0.024$, $t = 1.338$, $p > 0.05$).

Discussions

While the excitement surrounding supply chain digitalization continues to increase, there is a notable dearth of empirical research examining the precise mechanisms through which it can generate business value (Ye et al., 2022; Alvarenga, 2023). Considering that the expected outcomes of digital technology adoption are not always realized, it is crucial to acknowledge the lack of research concerning how supply chain digitalization specifically enhances firm performance. Accordingly, this study aims to address this issue and understand if and through what mechanisms supply chain digitalization results in supply chain robustness. Specifically, the study examined the mechanisms through which IT-enabled dynamic capabilities such as sensing, strategic learning, and strategic reconfiguration act as intervening variables to facilitate the effect of supply chain digitalization on supply chain robustness. The study draws from the resource orchestration theory to argue that supply chain digitalization can be orchestrated to develop dynamic capabilities which is required to achieve supply chain robustness.

The study examined the direct relationship between supply chain digitalization and organization sensing and learning capabilities. First, the findings of the study confirmed that supply chain digitalization had a direct relationship with organizational sensing capabilities ($\beta = 0.250$; $t = 4.447$; $p < 0.05$). This result aligns with the assertions in the extant literature (Mikalef et al., 2020). This finding suggests that the adoption and integration of digital technologies within the supply chain enhances the organization's ability to sense and monitor relevant information from internal and external sources. Digitalization enables organizations to gather real-time data on inventory levels, demand patterns, supplier performance, market trends, and other factors. This data availability and accessibility enable improved sensing capabilities by providing a comprehensive and up-to-date understanding of the supply chain environment. Organizations can leverage advanced analytics tools available in emerging digital supply chain technologies to harness and analyze data from various sources, leading to more accurate and timely sensing of supply chain dynamics.

Second, the results show that supply chain digitalization is positively related to organizational learning capabilities. These results lend support to previous studies that have observed that digital technologies may be leveraged to build organizational learning dynamic capabilities (Torres et al., 2018). By leveraging digital technologies, organizations gain access to a wealth of data and information that can be analyzed and interpreted to derive insights and knowledge. Digitalization facilitates the collection, storage, and analysis of large volumes of data, enabling organizations to identify patterns, trends, and relationships that were previously inaccessible or difficult to uncover. Improved learning capabilities enabled by supply chain digitalization foster a culture of continuous improvement, enabling organizations to identify areas for enhancement, refine strategies, and adapt to changing market conditions.

Further, the study examined the mediating roles of sensing and learning in the relationship between digitalization and organizational reconfiguration capabilities. The study found that while organizational learning capabilities mediate the relationship between supply chain digitalization and organizational reconfiguration capabilities, sensing capability does not. This suggests that the link between digitalization and supply chain robustness is not mediated by sensing capability. The mediating role of learning suggests that supply chain digitalization positively influences reconfiguration capabilities through its impact on learning. Digitalization enhances the organization's ability to learn from data, gain insights, and identify the need for reconfiguration. It also improves the organization's sensing capabilities, allowing them to monitor the supply chain environment, identify opportunities, and proactively respond to changes. These improved learning capabilities then facilitate the organization's reconfiguration efforts, enabling them to adapt and optimize their supply chain configurations in response to the dynamic business environment.

The study investigated the role of reconfiguration as a mediator in the relationship between supply chain digitalization, learning capability, sensing capability, and supply chain robustness. The results demonstrate that reconfiguration acts as a mediator between digitalization and supply chain robustness, as well as between learning capabilities and supply chain robustness. This suggests that digitalization has a positive influence on reconfiguration and learning capabilities, which, in turn, enhances the robustness of the supply chain. Digitalization grants organizations access to a vast amount of data and information, which can be analyzed and translated into valuable insights. By leveraging digital tools and analytics, organizations can improve their learning capabilities, enabling them to make data-driven decisions, continuously enhance processes, and optimize supply chain performance. This enhanced learning empowers organizations to identify areas for improvement, respond to market changes, and make strategic adjustments, ultimately bolstering the robustness of the healthcare supply chain. By utilizing digital technologies and enhancing their reconfiguration and learning capabilities, organizations can adapt their supply chain configurations, make well-informed decisions, and optimize performance. This, in turn, leads to a more robust supply chain.

Contrary to our initial hypothesis, we did not find evidence to support the direct mediating role of reconfiguration in the relationship between sensing and supply chain robustness. However, upon conducting a post hoc analysis, we uncovered an alternative pathway that sheds light on the underlying dynamics. Specifically, our results revealed that the impact of sensing capabilities on supply chain robustness is not solely mediated by reconfiguration. Instead, a sequential mediation pathway involving learning and reconfiguration emerged as a significant factor in understanding the relationship between sensing and supply chain robustness. Our findings suggest that sensing capabilities play a pivotal role in enabling supply chain robustness by first facilitating learning capabilities. Building on the insights gained through sensing, organizations with improved learning capabilities are better equipped to identify areas for improvement, refine strategies, and make informed decisions regarding reconfiguration. Learning from data analysis, insights generation, and knowledge sharing enables organizations to adapt their supply chain configurations to align with changing market conditions, customer demands, and other factors. This finding highlights the importance of understanding the sequential dynamics of sensing, learning, and reconfiguration in achieving a robust supply chain in the context of supply chain digitalization.

Conclusion

The study examined how supply chain digitalization may be leveraged to enable supply chain robustness. The study provides detailed expositions on the critical role of IT-enabled dynamic capabilities in enabling organizations to build robust supply chains from the use of supply chain digitalization. This study makes some vital theoretical contributions. First, this study is an attempt to contribute to the information systems and supply chain robustness literature and respond to a call for more empirical research on the topic (Ivanov, 2020; Ivanov and Dolgui, 2020). Again, this study empirically unfolds the crucial role of IT-enabled dynamic capabilities in achieving supply chain robustness when supply chain disruptions occur. By linking the ROT with IT-enabled dynamic capabilities, the study provides a clear view of how supply chain digitalization can be orchestrated to enable robust supply chains. Specifically, the result from this study highlights the critical unique roles of sensing, learning, and reconfiguration capability in allowing organizations to build robust supply chain systems from the use of supply chain digitalization.

The study also has important implications for organizations seeking to derive maximum benefit from supply chain digitalization implementation. First, the result of this study affirms that supply chain digitalization may be viewed as a strategic investment that enables organizations to business robust supply chains. By demonstrating the relationship between supply chain digitalization and supply chain robustness, the study provides empirical evidence that may serve as a basis for business managers to justify investments in supply chain digitalization. Again, the study highlights the critical role of sensing, strategic learning, and strategic reconfiguration in realizing the value of supply chain digitalization.

As with any research, this study is subject to some limitations that offer an opportunity for further research. First, the study adopted a cross-sectional design and investigates mainly the context of Ghanaian firms. Therefore, future studies may be undertaken in other countries to provide data regarding the similarities and/or differences with other contexts. Second, the study used cross-sectional data which was collected at one point in time. This limits the extent to which this study can establish causality relationships. Thus, future research should adopt a longitudinal approach in order to track the evolving effects of the relationship between supply chain digitalization, IT-enabled dynamic capabilities, and supply chain robustness.

REFERENCES

Agyei-Owusu, B., Amedofu, M. K., Asamoah, D., & Kumi, C. A. (2021, August). The effect of data-driven culture on customer development and firm performance: the role of supply chain information sharing and supply chain information quality. In Conference on e-Business, e-Services and e-Society (pp. 481-492). Cham: Springer International Publishing.

Alvarenga, M. Z., Oliveira, M. P. V. D., & Oliveira, T. A. G. F. D. (2023). The impact of using digital technologies on supply chain resilience and robustness: the role of memory under the COVID-19 outbreak. Supply Chain Management: An International Journal.

Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E: Logistics and Transportation Review, 114, 416–436.

Asamoah, D., Agyei-owusu, B., Andoh-baidoo, F. K., & Ayaburi, E. (2021). Inter-organizational systems use and supply chain performance: Mediating role of supply chain management capabilities. International Journal of Information Management, 58(December 2019), 102195.

Asiaei, K., Rezaee, Z., Bontis, N., Barani, O., & Sapiei, N. S. (2021). Knowledge assets, capabilities and performance measurement systems: a resource orchestration theory approach. Journal of Knowledge Management, 25(8), 1947–1976.

Badewi, A., Shehab, E., Zeng, J., & Mohamad, M. (2018). ERP benefits capability framework: orchestration theory perspective. Business Process Management Journal, 24(1), 266–294.

Brandon-Jones, E., Squire, B., Autry, C. W., & Petersen, K. J. (2014). A contingent resource-based perspective of supply chain resilience and robustness. Journal of Supply Chain Management, 50(3), 55-73.

Burin, A. R. G., Perez-Arostegui, M. N., & Llorens-Montes, J. (2020). Ambidexterity and IT competence can improve supply chain flexibility? A resource orchestration approach. Journal of Purchasing and Supply Management, 26(2), 100610.

Chadwick, C., Super, J. F., & Kwon, K. (2015). Resource orchestration in practice: CEO emphasis on SHRM, commitment-based HR systems, and firm performance. Strategic Management Journal, 36(3), 360-376.

Chesbrough, H. (2020). To recover faster from Covid-19, open up: Managerial implications from an open innovation perspective. *Industrial Marketing Management*, *88*, 410-413.

Chowdhury, M.M.H., Quaddus, M. (2017). Supply chain resilience: Conceptualization and scale development using dynamic capability theory. International Journal of Production Economics, 188, 185-204

Cui, L., Wu, H., Wu, L., Kumar, A., & Tan, K. H. (2022). Investigating the relationship between digital technologies, supply chain integration and firm resilience in the context of COVID 19. Annals of Operations Research, 1-29.

Davis-Sramek, B., Germain, R. and Krotov, K. (2015). Examining the process R&D investment-performance chain in supply chain operations: the effect of centralization", International Journal of Production Economics, Vol. 167, pp. 246-256.

Deepu, T. S., & Ravi, V. (2021). Supply chain digitalization : An integrated MCDM approach for inter-organizational information systems selection in an electronic supply chain. International Journal of Information Management Data Insights, 1(2), 100038.

Ehie, I., & Ferreira, L. M. D. (2019). Conceptual development of supply chain digitalization framework. IFAC-PapersOnLine, 52(13), 2338-2342.

El Baz, J., & Ruel, S. (2021). Can supply chain risk management practices mitigate the disruption impacts on supply chains' resilience and robustness? Evidence from an empirical survey in a COVID-19 outbreak era. International Journal of Production Economics, 233, 107972.

Essuman, D., Ataburo, H., Boso, N., Anin, E. K., & Appiah, L. O. (2023). In search of operational resilience: How and when improvisation matters. Journal of Business Logistics, 44, 300–322.

Galunic, D. C., & Rodan, S. (1998). Resource recombinations in the firm: Knowledge structures and the potential for Schumpeterian innovation. Strategic management journal, 19(12), 1193-1201.

Gelhard, C., von Delft, S., & Gudergan, S. P. (2016). Heterogeneity in dynamic capability configurations: Equifinality and strategic performance. Journal of Business Research, 69(11), 5272–5279.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). Sage Publications.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European business review, 31(1), 2-24.

Helfat, C. E., Finkelstein, S., Mitchell, W., Peteraf, M., Singh, H., Teece, D., & Winter, S. G. (2009). Dynamic capabilities: Understanding strategic change in organizations. John Wiley & Sons.

Hennelly, P. A., Srai, J. S., Graham, G., & Fosso Wamba, S. (2020). Rethinking supply chains in the age of digitalization. Production Planning & Control, 31(2-3), 93-95.

Koufteros, X., Verghese, A., & Lucianetti, L. (2014). The effect of performance measurement systems on firm performance: A cross-sectional and a longitudinal study. Journal of operations Management, 32(6), 313-336.

Kumi, C. A., Asamoah, D., & Agyei-Owusu, B. (2021). How Does IOS-Enabled Business Intelligence Enhance Supply Chain Performance?. In Responsible AI and Analytics for an Ethical and Inclusive Digitized Society: 20th IFIP WG 6.11 Conference on e-Business, e-Services and e-Society, I3E 2021, Galway, Ireland, September 1–3, 2021, Proceedings 20 (pp. 441-453). Springer International Publishing.

Liu, H., Wei, S., Ke, W., Wei, K. K., & Hua, Z. (2016). The configuration between supply chain integration and information technology competency: A resource orchestration perspective. Journal of Operations Management, 44, 13-29.

Mandal, S., & Dubey, R. K. (2021). Effect of inter-organizational systems appropriation in agility and resilience development: an empirical investigation. Benchmarking: An International Journal, 28(9), 2656–2681.

Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. Journal of Business Research, 70, 1-16.

Mikalef, P., Pateli, A., & van de Wetering, R. (2021). IT architecture flexibility and IT governance decentralisation as drivers of IT-enabled dynamic capabilities and competitive performance: The moderating effect of the external environment. European Journal of Information Systems, 30(5), 512-540.

Muneeb, D., Khong, K. W., Ennew, C., & Avvari, M. (2019). Building an integrated conceptual model of competitive learning capability: A strategic management perspective. Asia-Pacific Journal of Business Administration, 11(3), 267–287

Pavlou, P. A., & El Sawy, O. A. (2011). Understanding the elusive black box of dynamic capabilities. Decision sciences, 42(1), 239-273.

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. Journal of management information systems, 10(2), 75-105.

Qrunfleh, S., Vivek, S., Merz, R., & Mathivathanan, D. (2023). Mitigation themes in supply chain research during the COVID-19 pandemic: A systematic literature review. Benchmarking: An International Journal, 30(6), 1832-1849.

Richey Jr, R. G., & Davis-Sramek, B. (2021). Can we connect? Benefits and barriers to adoption of transformative technology in global supply chains. Journal of Business Logistics, 42(1), 4-6.

Rojo, A., Burin, G., Perez-arostegui, M. N., & Llorens-montes, J. (2020). Ambidexterity and IT competence can improve supply chain flexibility ? A resource orchestration approach. Journal of Purchasing and Supply Management, 26(2), 100610.

Ruel, S., & El Baz, J. (2021). Disaster readiness' influence on the impact of supply chain resilience and robustness on firms' financial performance: A COVID-19 empirical investigation. International Journal of Production Research, 61(8), 2594-2612.

Saarikko, T., Westergren, U. H., & Blomquist, T. (2020). Digital transformation: Five recommendations for the digitally conscious firm. Business Horizons, 63(6), 825-839.

Sarkis, J. (2020). Supply chain sustainability: learning from the COVID-19 pandemic. *International Journal of Operations & Production Management*, *41*(1), 63-73.

Simchi-Levi, D., Wang, H., & Wei, Y. (2018). Increasing supply chain robustness through process flexibility and inventory. Production and Operations Management, 27(8), 1476-1491.

Sirmon, D. G., Hitt, M. A., Ireland, R. D., & Gilbert, B. A. (2011). Resource orchestration to create competitive advantage: Breadth, depth, and life cycle effects. Journal of management, 37(5), 1390-1412

Teece, D. J. (2007). Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic management journal, 28(13), 1319-1350.

Teece, D. J. (2012). Dynamic capabilities: Routines versus entrepreneurial action. Journal of management studies, 49(8), 1395-1401.

Wong, C. Y., Wong, C. W., & Boon-itt, S. (2015). Integrating environmental into supply chain: A systematic literature review and theoretical framework. International Journal of Physical Distribution & Logistics Management, 45(1/2), 43–68.

Xie, X., Wu, Y., Palacios-Marqués, D., & Ribeiro-Navarrete, S. (2022). Business networks and organizational resilience capacity in the digital age during COVID-19: A perspective utilizing organizational information processing theory. Technological Forecasting and Social Change, 177, 121548.

Yang, L., Huo, B., Tian, M., & Han, Z. (2021). The impact of digitalization and inter-organizational technological activities on supplier opportunism: the moderating role of relational ties. International Journal of Operations & Production Management, 41(7), 1085-1118.

Ye, F., Liu, K., Li, L., Lai, K. H., Zhan, Y., & Kumar, A. (2022). Digital supply chain management in the COVID-19 crisis: An asset orchestration perspective. International Journal of Production Economics, 245, 108396.

Zahra, S. A., & George, G. (2002). Absorptive capacity: A review, reconceptualization, and extension. Academy of management review, 27(2), 185-203.

DECISION SCIENCES INSTITUTE

Buyer-Supplier Power Dynamics When Developing Innovative Sustainable Foods

Vahid Mirzabeiki
University of Surrey
Email: v.mirzabeiki@surrey.ac.uk

James Aitken
University of Surrey
Email: james.aitken@surrey.ac.uk

ABSTRACT

Through multiple case studies, this paper explores the changes in the buyer-supplier power asymmetry caused by the dynamism in the Innovative Sustainable Foods (ISF) market. Seven cases were studied to investigate the changing power landscape.

The findings of the study show that higher power by suppliers can lead to issues in the supply chains of buyers, making them vulnerable to the consequences of enhanced power of their suppliers.

The paper contributes to the literature on buyer-supplier relationship management, by highlighting the new trends in the food market and their impact on the power asymmetry between buyer and supplier.

KEYWORDS: Food supply chain, Procurement, Innovative sustainable foods, Power

DECISION SCIENCES INSTITUTE

Can an Innovative Pedagogical Approach Increase Elementary Statistics Course Achievement for Minority Students?

Candice L. Ridlon
University of Maryland, Eastern Shore
Email: cridlon@umes.edu

Camryn Walston
University of Maryland, Eastern Shore
Email: cawalston@umes.edu

Fabiola Beauvoir
University of Maryland, Eastern Shore
Email: fbeauvoir@umes.edu

ABSTRACT

In a two-year experiment, minority students taking Elementary Statistics at an HBCU were randomly selected from 300 undergraduates to be in a Control (n = 49) or Experimental (n = 43) group. Students had the same instructor, assignments, and assessments. The Control had traditional lectures, whereas the Experimental group learned via an innovative flipped learning pedagogy requiring them to submit pre-class notes and participate in real-life group problem-solving activities. Analysis yielded no significant difference between the group's final grades (p = 0.2494), but the flipped learning innovations correlated positively with final grades, while concurrently decreasing math anxiety and enhancing self-confidence.

KEYWORDS: Innovative classroom instruction, Flipped learning, Educational decision-making, Minorities, Math anxiety

INTRODUCTION

Students at Historically Black Colleges and Universities (HBCUs) in the United States tend to arrive on campus with significant skill deficits in mathematics. Their past failures in mathematics result in low self-confidence and anxiety towards any course labeled “math.” Because college-level statistics courses are generally housed in the Mathematics Department, many minority students are predisposed to believe they are incapable of passing such a class. Then, to prove their point, these students guarantee their own failure by exhibiting behaviors that ensure this becomes a self-fulling prophecy.

Low-achieving minority students are often subjected to instruction models aimed at basic remediation rather than being given opportunities to develop critical thinking skills. Their college-level statistics curriculum tends to focus on memorization of shallow, unconnected methods that can be “simply” associated with definitions, tables, and formulas. This approach increases student anxiety towards the subject -- and worse yet, does not foster any understanding of the powerful concepts inherent in decision sciences.

However, innovative instructional approaches that employ real-world problems and data driven activities (e.g. problem-centered learning, flipped learning, active learning) have the potential to help low-achieving minorities see their success and failure as self-controllable rather than attributed to external factors (like their instructor or the subject matter). This implies that universities offering undergraduate courses in subjects related to the decision sciences (especially those with high minority enrollment) need to consider using the kind of pedagogy that cultivates a diverse, competent population of professionals who know how to analyze and interpret data in their respective fields.

This paper explores the use of an innovative pedagogy called “flipped learning” in a college-level introductory statistics course at an HBCU. We will explore the impact of this instructional model while considering the underlying themes of minority education, math anxiety, and student self-confidence. From the evidence presented, we will decide if this pedagogical approach has the potential to support minority students in the development of higher-order reasoning skills. Our main research question is: “Does the flipped learning pedagogy have an effect on minority students’ achievement in a college-level Elementary Statistics course?”
We also contemplated three sub-questions related to achievement that would inform our decision-making process and lead us towards a well-considered answer to our main research question:

1) Does the flipped learning pedagogy have an effect on minority students’ math anxiety and self-confidence, both of which influence achievement?
2) Is there any identifiable difference in Black males’ achievement when they participate in the flipped learning pedagogy?
3) Does the attitude that the below-average performance is not only sufficient, but also acceptable, in Non-STEM disciplines affect student achievement, making it difficult to cultivate a diverse, competent population of professionals?

LITERATURE REVIEW

We now provide a description of some of the major underlying themes related to the teaching and learning of Elementary Statistics, one of the introductory courses falling under the umbrella of Decision Sciences. We also describe the challenges facing the minority student populations who are at the heart of this study.

Minority Education, Mathematics, and Statistics

Students who enroll at Historically Black Colleges and Universities (HBCUs) in the United States generally possess significant skill deficits. Average admissions requirements across all 105 HBCUs (public, private, and two-year colleges) are a 2.5 GPA – and if they require admission test scores at all, then the average is an 18 on the ACT or a 905 on the SAT (Toldson & McGee, 2014). To remediate this widespread shortfall in mathematics skills, HBCUs typically require that freshmen (usually about 90% of them) take one prerequisite math course (generally below college-level, and that course does not count towards graduation). The HBCU in this study shares these characteristics and subscribes to the 1-course remediation strategy. The majority of universities (including HBCUs) house statistics courses in the Department of Mathematics because statistics is a branch of applied mathematics. This makes sense because the theories behind statistics rely heavily on differential and integral calculus, linear algebra, and probability theory (Chapelow, 2023). Thus, the prefix for an Elementary Statistics course is usually MAT or MATH. For students, this translates into statistics = math. Because they view stats as just another math course, their expectations for success in statistics tends to be closely associated with their previous experiences in math.

In harmony with this math-statistics relationship, all universities require the prerequisite of some college-level mathematics course, usually College Algebra, before enrolling in Elementary Statistics. However, many HBCUs (such as ours) allow non-STEM majors like Rehabilitation Psychology, Special Education, or Exercise Science to take statistics after “passing” only one low-level course similar to high school Algebra 1, which is considerably less robust than College Algebra. These same majors also accept a “D” in Elementary Statistics to count towards graduation. These facts underscore the reality that for most students, only minimal preparation is required before enrolling in Elementary Statistics - and that below-average performance is not only sufficient, but also acceptable. This attitude is contrary to the belief that the world needs a diverse, competent population of professionals who know how to analyze and interpret data in their respective fields.

Regardless minority status, gender, or other classification, college students uniformly report they feel disconnected from math and that they struggle to pass mathematics courses (Pringle et al., 2022). What is more, mathematics anxiety adds an additional challenge for many four-year university students. Symptoms of math anxiety include negative self-talk, feeling overwhelmed, lack of motivation to do math, relying on memorization rather than understanding, and not being able to recall concepts on a test when the student knew them before the exam (called “going blank”) (Sokolowski & Ansari, 2017). Although statistics instructors are familiar with these behaviors and such conduct can be objectively measured, the teacher may mistakenly believe the student is immature or lazy rather than recognizing these as hallmarks of the condition. Additionally, without medical expertise it is unrealistic to expect an instructor to verify the observable physical symptoms of math anxiety, such as increased heart rate, sweaty palms, and lightheadedness (Luttenberger et al., 2018) that allows them to adjust their teaching style to accommodate the problem.

Math anxiety can pose a barrier to successful completion of the mathematics courses necessary to graduate. In fact, a meta-analysis of many international studies by Luttenberger et al. (2018) found that “math anxiety correlates negatively with math achievement … and this relationship remained stable after several assessment periods.” Another study found that approximately 25% of all four-year college students taking mathematics courses struggle with moderate to high math anxiety (Beilock & Willingham, 2014). The percentage of minority students may be even greater because the high incidence of their past failures amplify the potential for anxiety. It is important to note that math anxiety is analogous to statistics anxiety, because as mentioned previously, from a student’s viewpoint stats is math.

The Impact of Self-Confidence on Success for Minorities

Low self-confidence due to past failures, added to a cultural pressure to appear self-sufficient (because asking for help means you are weak), often causes many minority students to believe they are incapable of succeeding in math. Rather than trying to achieve, they fall into a cycle of continued failure by exhibiting behaviors that guarantee they will fail. Educators have identified several factors contributing to this unproductive chain of events. As early as 1997, Steele found “minorities tend to develop a perception of their ability, intellectual performance, and identity of self-based on stereotypes.” According to Johnson (2000), that stereotype is that they will perform poorly in math.

The narratives about Black males in math education are generally negative. Commonly held destructive stereotypes of low achievement are articulated both in terms of teachers’ expectations and in how Black males behave in response to those beliefs. They tend to self-limit their engagement in school and may even suppress the possibility of positive academic outcomes. Black males are known to engage in well documented self-defeating behaviors such as “showing off what they were expected to learn and be like – i.e. talking and ‘holding the floor’- rather than trying to develop the quality, deeper conceptualized understandings of mathematics” that their teachers were trying to foster during class (Martin, 2000; Murrell, 1995). In any case, harmful narratives eventually influence how Black males make sense of their ability to perform in math (Jackson et al., 2021).

Regardless of ethnicity, considerable research has documented the empirical link between a student’s academic identity (the connection between one’s personal identity and one’s role as a learner within the academic domain) and their resulting academic achievement (Jackson et al., 2021). In addition, the math achievement gap between Black males and other groups (even Black females) is not only published by the academic community, but is readily observable by teachers and students throughout the K-12 years. Knowing that others generally outperform you in math creates a mindset that both expects and accepts failure. This attitude has influenced life opportunities for Black men and has perpetuated the cycle of generational inequality (Reeves & Halikias, 2017). We see the result in HBCUs, where close to 35% of minority students have difficulties moving forward in their major program and eventually change majors purely because they were unable to pass the required math courses (Khasawneh et al., 2021; Ridlon, 2013).

Many studies have made the connection between a student’s belief in their ability to do math and their subsequent performance due to that belief. A meta-analysis of 40 research publications in national and international journals found that self-confidence and success in math are highly correlated. The analysis “indicate that there is a positive and significant relationship between self-efficacy and students' mathematics learning ability. This means that the higher the self-efficacy, the higher the mathematical ability, and the lower the self-efficacy, the lower the mathematical ability” (Muhtadi et al., 2022).

Pedagogical Approaches to Teaching Math and Statistics

When external measures like past low achievement and poor standardized test scores merge with self-inflicted or cultural expectations of failure, the customary educational response is to subject minority students to basic skills instruction in a well-intentioned effort to make statistics “accessible.” However, regardless of the subject matter, memorization of shallow, unconnected steps is unsustainable and that “knowledge” is too soon forgotten. Therefore, a college-level statistics instruction model that focuses on “simplifying” concepts so that they are reduced to a series of definitions, tables, and formulas is self-defeating. Rather than comprehending the big ideas of statistics and improving their decision-making capabilities, student anxiety is reinforced when they find it nearly impossible to memorize the unrelated steps and calculations they “saw” over a four-month semester. Any opportunity for understanding the powerful concepts inherent in the decision sciences is squelched because the curriculum never exposed them to real-world problem-solving activities that are vital to developing higher-order reasoning skills. Students who learn statistics with this kind of fundamental approach are apt to lose any confidence they might have had in their own ability to think mathematically (Ridlon, 2009, 2013).

However, working with minorities, Mulat and Arcavi (2009) reported that success in mathematics learning can shift by exposure to new ways of thinking about math … low-achieving minorities can see their success and failure as attributable to self-controllable (e.g. fostered self-regulation) rather than as a result of uncontrollable external factors (e.g. their instructor or the subject matter). This realization helped students take responsibility for their own learning and invest in their studies. Once they saw the connection, their anxiety decreased and their motivation to learn increased.

Innovative instructional approaches that engage students' interest by using real-world problems and data-driven activities (e.g. problem-centered learning, flipped learning, active learning) hold the potential to transform a student's relationship with statistics. "Active learning" is not just the latest academic fad – on the contrary, active learning is a well-tested approach (Mills, 2012). Swigart and Murrell (2001) found that active instructional strategies resulted in greater gains in learning. They point out that these strategies are especially important for first-generation college students in the absence of role models who can help these students understand their own responsibility for learning and development – typical of the minority student body at an HBCU.

For the past decade, a pedagogy called "flipped learning" has been widely touted for its ability to increase academic achievement while concurrently decreasing math anxiety and bolstering students' self-confidence at the college level. Bredow and her colleagues (2021) performed a meta-analysis of data from 317 studies and found that, "the effects of flipped learning in higher education produced positive gains across all three learning domains… with significant advantages of flipped over lecture-based instruction for seven out of eight outcomes." Of critical importance for the decision sciences (where problem solving and decision-making are essential skills), flipped classroom models have been shown to have a statistically significant influence on students' use of higher-order cognitive skills (Jung et al., 2022). Also, relevant to decision sciences, Albert and Beatty (2014) used the flipped classroom model in an introduction to management course at a highly diverse, urban, Business–accredited university and found that compared to lecture classes taught by the same instructor, grades on exams were higher.

To examine the effect of the flipped learning pedagogy in Elementary Statistics, this paper analyzes data collected over a two-year period at an HBCU. We will discuss the impact of this instructional approach on student achievement, math anxiety, and self-confidence. We will also segregate and interpret data collected on Black males who participated in the study.

THEORETICAL MODEL: THE FLIPPED LEARNING PEDAGOGY

Then innovative instructional model called "flipped learning" we used in our HBCU's Elementary Statistics course to investigate our research question and its three sub-questions has three major "tenants." We describe our execution of these three tenants in the table below.

Table 1: The Flipped Learning Model in Elementary Statistics

Tenants of Flipped Learning	Implementation of the Tenants
1) Provide the opportunity for students to gain exposure to course content prior to class.	1) Twice per week, students prepared for class by submitting pre-class hand-written summaries on basic information (posted on *PowerPoint* slides) and by taking notes on short videos. The incentive for completing these "Chapter Notes" (called "Notes") was straightforward: Notes counted 15% of the overall grade.
2) Provide an objective mechanism to assess student understanding.	2) All assignments (homework, quizzes, and exams) were objectively scored by a curriculum website (*Hawkes Learning*). Students could check their progress at any time. We provided a Teaching Assistant, who tutored students on-demand and free of charge. The TA attended class on Class

	Activity days to work with groups, graded students' Notes, and communicated with the instructor so that we could objectively assess individual student understanding. To encourage students to take sufficient time to understand course material and work with the TA, their homework could be submitted up to 5 days late with gradually increasing late penalty. Formative and summative assessments consisted of an online quiz or exam that occurred every week, and these assessments could be taken twice within a 7-hour window. The retake option allowed students time to study material after their first attempt. Only the best score was recorded in the gradebook.
3) Provide a clear connection between in-class and out-of-class activities.	3) Students completed weekly "in-class" group problem-solving "Class Activity" lasting approximately 50 minutes. These activity worksheets were comprised of data-driven real-world problems that required application of the material learned on the *PowerPoint* slides presented "in-class" during lectures. The problems on the Class Activity were aligned with the Notes and videos students completed out-of-class that same week, as well as containing the same types of questions found on the "out-of-class" homework, weekly quizzes, and tests. The TA conducted weekly out-of-class Study Sessions to work problems similar to those found on that week's assessment.

One of the most striking differences between flipped learning model and the traditional lecture approach is that this instructional innovation requires students to take responsibility for their own learning and personally invest in their studies. A big percentage (15%) of the student's grade depended on their taking twice-per-week "Chapter Notes" (called Notes hereafter) on the *PowerPoint* slides we posted on our Learning Management System (*Canvas*) and watching/taking notes on accompanying videos <u>before</u> they came to class. According to Doo (2022), the advantages of pre-learning sessions in the flipped model is that this activity facilitates the understanding of lectures and provides review opportunities for learning (students can go back and "read again" in real time as needed, plus they know what questions they want to ask prior to coming to class). Other benefits are the flexible learning time (students can prepare Notes when it suits their schedule) and the opportunity for individualized learning. An example one typical Notes assignment is found in Appendix 1.

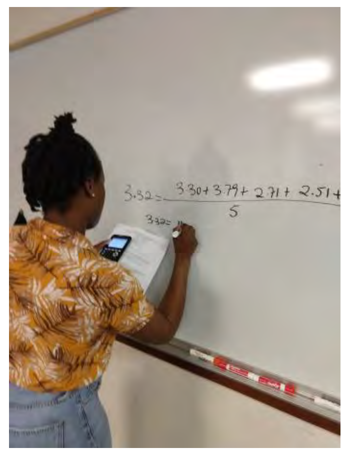
The Peer Tutoring (Teaching Assistant) Component of Flipped Learning

Peer tutoring has many advantages, and flipped learning proponents suggest choosing some means of incorporating tutoring to facilitate learning. Recommendations range from having a tutor actively assisting in class every day (robust support) to offering on-demand virtual coaching (minimal support). We chose to employ one Teaching Assistant (TA) per section of Elementary Statistics to scaffold instruction throughout the semester (as indicated in the chart above: Implementation of the Tenants 2 and 3). This individual was an undergraduate student who had previously succeeded in his or her own Elementary Statistics course. In-person tutoring was provided free-of-charge and made available upon student request via text or email. The TA's other responsibilities included:

- Attending class on Class Activity days, moving around to ask/answer questions.
- Grading Notes assignments (handwritten notes on *PowerPoint* slides, comments on the required videos, and Activity worksheets). The TA entered scores in the gradebook.

- Conducting weekly study sessions in preparation for the online quizzes and tests. The TA created "practice versions" of each assessment and students worked together to complete these handouts.
- When a student was struggling, aside from the academic coaching one would expect, the TA provided encouragement and support. They could discourage negative self-talk and remediate the symptoms of math anxiety by recommending proven coping mechanisms, like taking a deep breath or recognizing past success. These suggestions were less threatening when they came from a peer rather than an instructor.
- Working in a one-on-one situation, the TA could objectively assess student effort and content understanding so that individual accommodations could be made when warranted (such as extending due dates on homework assignments without a late penalty because a student was diligently working, not merely disregarding assignments)

Picture 1:
TA working with students during
Class Activity



Picture 2:
TA modeling problem-solving on
the board

The Classroom Culture in Flipped Learning

Our flipped learning classroom looked quite different from those in a traditional lecture. First, the students sat in groups of 4 to 5 clustered around tables rather than sitting separately in rows. Students worked in collaborative groups daily so the classroom arrangement needed to foster communication. Second, in the flipped Statistics classroom, technology was used daily. Students accessed the internet and software (e.g. IBM's *SPSS*) to investigate real-world problems. They regularly referred to the *PowerPoint* slides posted on *Canvas*. Thus, computers were situated in the center of clustered tables, or students used their own machines. Finally, a calculator with statistics functions was mandatory. (We provided loaner TI-83s if needed.) By the third week of the semester, we used calculators every day during class time.

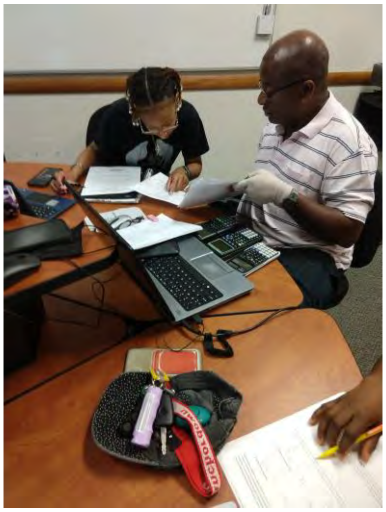
Because students turned in handwritten summaries of basic terms and watched relevant videos <u>before</u> class, time was now devoted to discussing big ideas, modeling different problem-solving strategies, and interpreting findings in statistical terms as well as in everyday language. We used *PowerPoint* slides to display a variety of real-life situations. The slides also captured TI graphing calculator screens to illustrate desired keystrokes and functions. Therefore, instead of the passive note-taking behavior that dominates the usual college lecture, flipped learning students were actively working together and having lively discussions to defend their logic in reaching a conclusion. Three typical examples of *PowerPoint* slides (obviously without line-by-line animation) are provided in Appendix 2.

Another difference in the classroom culture was the instructor's behavior. Instead of standing in front of the class to lecture, once a problem was introduced, the instructor circulated around the room to assist individuals or groups in their problem-solving activity. Rather than students sitting quietly and listening, the flipped classroom was filled with animated conversation between students, instructor, and TA. The room could be noisy as students enthusiastically rejoiced in their "a-ha" moments or argued over interpretations. All present observed the development of students' self-efficacy when they found that they were able to think mathematically.

The Active Learning Strategy in Flipped Learning

During the in-class stage of flipped learning, the major learning strategy should be problem-based learning (Chen et al., 2022). In our study, two-thirds of the class time was dedicated to the instructor modeling problem-solving strategies and decision-making, and afterward giving students the chance to practice those same schemes at their clustered tables. The remaining one-third of class time (or one day per week) was dedicated to group problem-solving Class Activities. These Activity worksheets contained new, unique problems, and accounted for 40-60% of the students' Notes grade. Class Activities were data-driven and relevant to student interests. Each worksheet asked students to interpret findings and make appropriate decisions and/or conclusions**.** Three examples of Class Activity worksheets are presented in Appendix 3.

Aside from the real-life problem solving skills students were developing in Activities, we had a secondary goal in mind. We wanted to facilitate an atmosphere of supported (yet independent) critical thinking, so that students did not expect the teacher to be the sole source of answers. Instead, by working with their peers, we hoped they would develop confidence in their own ability to solve problems and make decisions. As Mulat and Arcavi (2009) proposed, we believed it was possible for low-achieving minorities to see their success and failure as attributable to self-controllable (e.g. fostered self-regulation) rather than because of uncontrollable external factors (e.g. their instructor or the subject matter). We wished them to take responsibility for their own learning and invest in their studies.


Pictures 3, 4, and 5:
Students working together on the weekly group problem-solving Class Activity

Theory of Decision Sciences as Applied to the Flipped Learning Pedagogy

Decision Sciences is the study of decision making, with particular focus on solving real-world problems. The stated aim of the Decision Sciences Institute (DSI) is "to improve the judgement and decision-making capabilities of individuals, groups, and organizations across all areas of human enterprise." (DSI Homepage, 2023). Because all human endeavors could benefit by making better, efficient, and worthy decisions, Decision Science has become an interdisciplinary topic studied in many fields, each with varied approaches and underlying theories.

Educational decision-making is usually considered to fall under the category of Behavioral Decision Theory, of which there are two interrelated facets, normative and descriptive. According to Slovic, Fischhoff, and Lichtenstein (1977), "the normative theory is concerned with prescribing courses of action that conform most closely to the decision maker's beliefs and values. Describing these beliefs and values and the manner which individuals incorporate into their decisions is the aim of descriptive decision theory." Simplified, one could state that educational decision-making involves value judgements.

One might rightfully ask: Does any theory of Decision Science define the components of "optimal" value judgements for educational decision-making? Value judgments are notorious for being subjective rather than objective. The laws of applied probability cannot guarantee accurate predictions of outcomes for student behavior in the classroom because they are not designed to select "mathematically" the best models of instruction with estimates of success or failure. However, we can analyze data and obtain descriptive measures of success/failure, define and then reject/accept hypotheses, and look for correlations between various teaching innovations and their impact on learning. Nevertheless, as decision-makers aim for improvements in education, they need standards that tell them what counts as an improvement.

Harry Brighouse, a British philosopher, and his colleagues suggested a framework for Behavioral Decision Theory in education in their book *Educational Goods* (Brighouse et al., 2018). Their basic tenant is that improvement in education should focus on values that develop students' "capacities" and allow them to flourish (e.g. lead happy, satisfying lives). They described six of capacities, four of which are particularly relevant to our study. These capacities are described in Table 2, linking each construct to the flipped learning pedagogy.

Table 2: *Educational Goods* Values associated with Behavioral Decision Theory

Capacity	Description of Capacity	Development of Capacities in the Flipped Learning Pedagogy
Personal Autonomy	Ability to make and act on well-informed and well thought-out judgments	1. Work with real-life problem-solving situations 2. Learn how to collect and analyze sample and population data (ability assessed in homework/ exams) 3. Write and test hypotheses (ability assessed in homework/ exams) 4. Make decisions, draw conclusions
Democratic Competence	In a democratic society, citizens benefit from the ability to use their political institutions both to press their	1. Critical Thinking Standards explicitly taught and assignments using them are assessed

	own interests and to give due weight to the legitimate interests of others	2. Learn to use/interpret databases from CDC and other federal and international organizations
Economic Productivity	Ability to participate effectively in the economy by 1.enhancing cognitive skills, and 2. encouraging the disposition to work (which is also in the interest of the broader society)	1. Higher achievement = better job opportunities 2. Teaches students to work because they must take responsibility for their own learning and invest in their studies
Treat Others as Equals	Having respect for the basic dignity of persons underlies the idea that everybody has the same basic human rights, so must the ability to think of all people as fundamentally equal in moral status	Working in groups means: 1. Listening to others 2. Cooperating with those different from self to reach a goal 3. Valuing diverse opinions

Furthermore, Brighouse (2018) suggested a four-step process (labeled Step 1 to 4 below) for successful educational decision-making once the valued capacities have been recognized. Considering our research question and its three sub-questions re the flipped learning pedagogy, we aligned values and decision-making to the Brighouse Theory value steps, as described:

1) Step1: Identify pertinent values
    - *One of the tracks in Decision Sciences is Innovative Education. Elementary Statistics is a foundational course for developing the basic skills that are valued in decision-making, so it is critical students understand the concepts taught in this course. We must find innovative ways to teach and learn Stats.*
    - *We are interested in combating math anxiety, boosting self-confidence, and increasing achievement for minority students.*
    - *We believe all students can learn statistics and have high expectations of success, regardless of gender, ethnicity, or other diversity classification.*
    - *We do not think that earning a "D" in Elementary Statistics is sufficient or acceptable for any college major considering the emphasis of evidence-based data-driven decisions that must be made in any current field of human endeavor.*
    - *We value a pedagogy that cultivates a diverse, competent population of professionals who know how to analyze and interpret data in their respective fields.*
2) Step 2: Identify key decisions
    - *Will instructors at our HBCU will make the commitment to participate in workshops and learn the tenants of the flipped learning model, and then be willing to devote the considerable amount of time needed to implement it?*
    - *The Math Dept. at our HBCU must decide to support the experiment by providing appropriate facilities (classrooms with group seating potential, computers, etc.)*
    - *Must find funding for TAs and loaner calculators if going to implement.*
    - *Identify available funding to replicate in future semesters (if none, abandon?)*
    - *Decide questions to ask students on a survey given at end of semester. (As this is qualitative data, may use these responses to triangulate quantitative data and/or explain quantitative findings.)*
    - *Decide: What evidence supports that the flipped learning pedagogy is successful for minorities? Final grades alone? Do other positive outcomes weigh as strongly, or is increased achievement the only outcome of value?*

3) Step 3: Use evidence to evaluate options in light of pertinent values
   - *To gather evidence, design an experiment so pre- and post-implementation data can be collected. Does evidence support values in Step 1?*
   - *If quantitative data does not show an increase in achievement, does other valuable evidence on math anxiety, or the qualitative evidence obtained from student surveys, balance the option in favor of choosing flipped learning?*
   - *We believe it would be beneficial if all majors at our HBCU required a minimum "C" in the course is needed to count towards graduation.*
4) Step 4: Weigh values and evidence to make decisions
   - *If the results (after analysis) prove effective at increasing achievement Elementary Statistics for minorities, we have decided to champion the flipped pedagogy at our HBCU and by publishing results.*
   - *We hope to find new insights into the pedagogy of teaching and learning Elementary Statistics. If found, the Math Dept. may decide to adopt the flipped learning for teaching in all sections of the course.*
   - *Of special concern to an HBCU, we desire to use evidence-based data obtained to inform our community and make decisions that support our Black male population in their pursuit of STEM careers that require success in mathematics coursework.*

METHODS

In this two-year long experiment, two groups of minority students at an HBCU were selected from the overall population of more than 300 students who enrolled in Elementary Statistics. Students randomly self-enrolled in sections of this course offered at a variety of times of day. Sections of the course taught by the same instructor were chosen and students in them were designated as either the Control (C, n = 49) or Experimental (E, n = 43) group. Both groups of students attended a 3-day-a-week 50-minute lecture period.

Similarities in Curriculum of the Two Groups

The C and E groups shared these curriculum components:

- The same instructor taught the class.
- They used the same textbook and online assignment site: Warren, Denley, Atchley (2020). *Beginning Statistics. Hawkes Learning*: Mount Pleasant, SC.
- Identical *PowerPoint* slides guided classroom lectures and were posted on *Canvas*
- They had identical homework assignments posted on the *Hawkes* site. This program is based on "mastery learning" and had three components: (1) Read the book; (2) Practice: contains video examples, an "Explain the Error" button, and tutorials showing solution steps; and (3) Certify where the student complete the actual homework and scores are recorded. When a student misses a preset percentage of questions in Certify (usually 20% wrong), the system sends them to Practice before allowing re-entry to the Certify phase. Unlimited attempts were allowed to Certify. Late penalties are imposed if work not completed by due date (graduated 10% off per day).
- The same assessments were taken online at *Hawkes* during an open 7-hour window.
  - Ten Quizzes: approximately 20 questions, two attempts permitted, best score entered into gradebook, each had 75-minute time limit. (10% of grade)

  - Four Pretests: ranged 46 - 81 questions, unlimited attempts permitted, no time limit, best score entered into gradebook, must earn 70% on the Pretest for the associated Test to be made available (10% of grade)
  - Three Tests: ranged 29 - 42 questions, two attempts permitted, best score entered into gradebook, each had 75-minute time limit. (30% of grade)
  - One Final Exam: 54 questions, one attempt, two-hour time limit. (20% of grade)
- A Study Session was offered from 7 – 8 PM every Wednesday to both groups. The TA created “practice versions” of each upcoming assessment that closely matched what students would be asked, and they worked together to complete these handouts.

Differences in Curriculum of the Two Groups

The differences in the curriculum and instruction method of the C and E groups were:

- The E Group learned via the flipped learning model, which required them to complete two pre-class activities: (1) compiling hand-written notes on basic information from the *PowerPoint* slides, and (2) writing summaries of short videos. (15% of grade)
- The instructor lectured in the C class by animating each question and its answer on the *PowerPoint* slides while students took notes. They had to copy the basic information as it was shown. In the E group, the instructor skipped basic information on the *PowerPoint* slides, went directly to the question scenarios, and then students were required to participate by giving possible answers <u>before</u> those solutions were revealed.
- The C group was exposed to peer-group problem-solving tasks during class (limited to less than 20 minutes per week). Only the student presenting was graded.
- The E group completed a weekly in-class group problem-solving activity lasting up to 50 minutes. (Score included in Notes grade)

Data Collected on Both Groups and Type of Analyses Performed

Evidence-based data from both the C and E groups was collected in the form of scores on homework, quizzes, Pretests, Tests, and the Final Exam. The grades on E groups Chapter Notes was also available in the course gradebooks. We were privy to student demographics in terms of self-identified ethnic group membership and gender, so we were able to use those classifications to explore our three research sub-questions and their relationship to the main question. The following statistical analyses were performed:

- Descriptive measures (mean, median, range, standard deviation), and Boxplots of both group’s final grades in the course
- p-values using the *t*-distribution to see if the components of flipped learning (Notes and Class Activities) made a significant difference in E group’s final grades in the course
- Pearson coefficient r and correlation coefficient r^2 for the E group Notes vs. final grades
- Scatter plot with regression model to identify relationships between the flipped learning component called Notes and final grades
- Data analysis of students in both the C and E groups comparing the proportions of students who did not attempt any more independent work (homework) prior to the end of the semester (which we called becoming “Frozen”), as well as at what point during the 14-week semester they stopped working
- Analysis of data from the three course tests showing the number of questions students in both the C and E groups left unanswered
- We singled out the performance of Black males to examine any differences between their behavior and others.

RESULTS AND DISCUSSION

Main Research Question: Does the flipped learning pedagogy have an effect on minority students' achievement in a college-level Elementary Statistics course?

Descriptive measures (mean, median, range, standard deviation) from the numerical data on the final course grades earned by the C and E groups are shown in the Table 3 below:

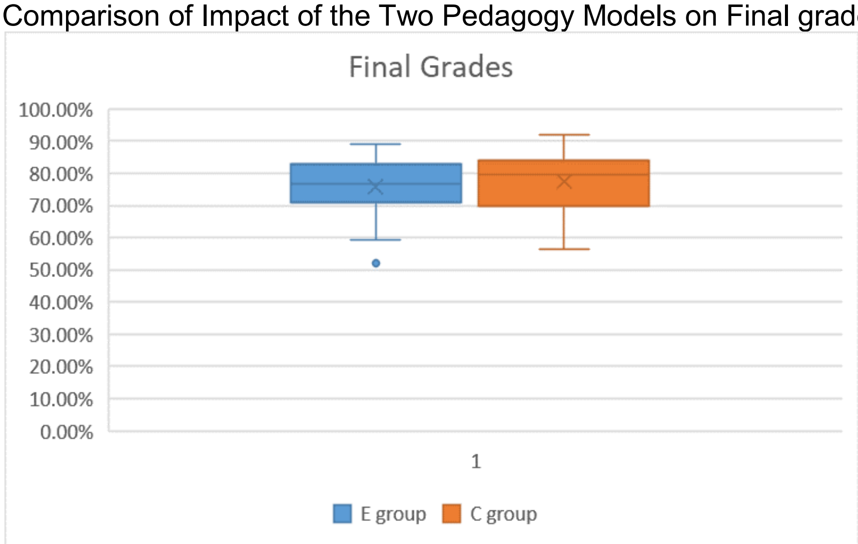
Table 3: Descriptive Measure of Final Grades in Elementary Statistics

	Mean	Median	Range	Standard Dev.
Control (n = 49)	77.41%	79.63%	35.57%	0.09116
Experimental (n = 43)	75.86%	76.66%	37.11%	0.089074

These descriptive measures yield an important observation. The mean and median final grades in the course of both the groups were very close, regardless of the instruction model used (traditional lecture or flipped learning). In fact, all descriptive measures were extremely close, including the standard deviation, implying that the samples might be representative of a normal distribution for the whole minority population of students taking Elementary Statistics and taught by this same instructor at this HBCU.

The final grades of the C and E groups are shown in boxplots below. This graph again illustrates the close outcome: there is little difference in achievement regardless of the pedagogy.

Graph 1: Comparison of Impact of the Two Pedagogy Models on Final grades



Note the singular dot E-group's final grades' boxplot. This point represents the one student out of 43 in the E group who stopped coming to class at 8 weeks, and did not attend the Final Exam. His behavior resulted in a very low F, represented as an outlier dot. It is interesting to note that this student was a Black male, and the only one out of the total 92 in the study who exhibited terminal absenteeism.

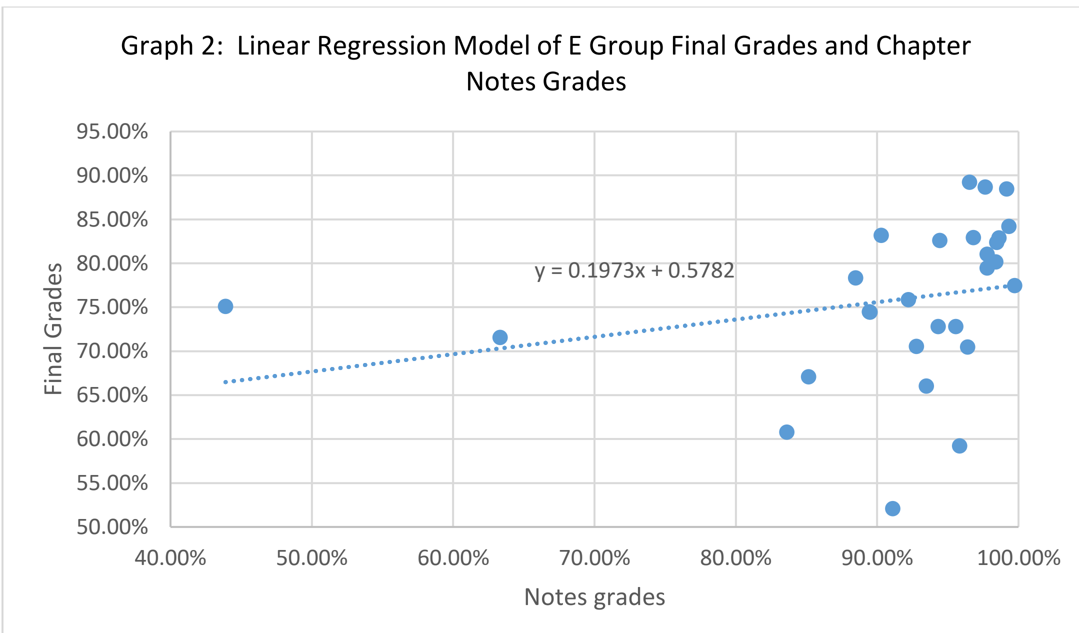
To statistically determine if there was any significant difference between the two groups' final grades, we performed a hypothesis test with μ_1 = C's mean final grade and μ_2 = E's mean final grade. The number of students in each sample group taking Elementary Statistics was greater than 30 (Control n = 49 and Experimental n = 43), and the descriptive analysis supported the assumption that population variances for final grades were equal for all students enrolled in the course. As the samples of students were randomly selected (students randomly enrolled into any section of Elementary Statistics), it was possible to use Central Limit Theorem. The two samples were independent. However, the population standard deviation σ was unknown; we only had the sample standard deviation $\bar{\varsigma}$. Thus, we were able to use the student t-distribution. We wanted to know if flipped learning pedagogy had a positive effect on minority students' achievement (did the pedagogy increase Final Grades?), so we tested the alternative hypothesis that $\boldsymbol{\mu_2} > \boldsymbol{\mu_1}$ with a significance level of $\alpha = 0.05$ (5%). The p-value obtained was 0.2494. Because $p > \alpha$ (0.2494 > 0.05), this indicated that there was not a significant difference between the mean of the two groups' final grades. Thus, we decided not to reject the null hypothesis stating there was no difference in the mean final grades (or $\boldsymbol{\mu_2} = \boldsymbol{\mu_1}$). It appeared that the flipped learning pedagogy did not have a substantial effect on minority students' achievement (as exhibited by final grades) in a college-level Elementary Statistics course.

In searching for an explanation, we believe the similarity in final grades might be attributed to the preparation students had for the high-stakes assessments in the course. As mentioned previously, students in both groups were required to earn a minimum score of 70% on four Pretests before they were able to take an exam, regardless of the instructional approach. These review assignments, coupled with weekly study sessions that examined the same kinds of questions asked on subsequent assessments, apparently did an excellent job of coaching students in the skills they needed to succeed. As the Pretests, Tests, and the Final Exam accounted for a major proportion (60%) of both group's final grade in the course, the impact of the flipped learning pedagogy with its innovations of Notes and Class Activities had a minimal influence on that particular outcome.

We were not alone in finding that flipped learning had a disappointing impact on students' achievement. Although we only compared two pedagogies (flipped and traditional), research by Burke and Fedorek (2017) found similarly disheartening outcomes when examining three approaches to teaching the same university criminology course. At the start of their study, they hypothesized that students in the flipped class would report higher levels of engagement compared to students in traditional or online courses, but instead they found engagement was surprisingly lower. Additionally, "it was thought that students in the flipped classroom would 'gain' more skills … but this was simply not true. In fact, students in the flipped classroom registered a much lower understanding of diverse people and/or diverse backgrounds" as well as having fewer critical thinking skills when compared with students in the traditional classroom, contrary to much of the literature. They attributed flipped learning's minimal influence on final outcomes (e.g. grades, enhanced engagement) to several limitations, such as: students may not know how to work effectively in groups, the model relies on students' agency to take an active role outside the classroom (which may not happen), and students' might have had an adverse reaction to drastic changes in pedagogy after being accustomed to years of lectures (Burke & Fedorek, 2017).

Fadol et al. (2018) also compared the three pedagogies (flipped, online, and traditional) in a college-level management course that consisted of females only. Unlike our study, in terms of final grades flipped learning performed the best, followed by online with traditional in last place. However, improved outcomes were associated with students' use of online materials, a constraint explicitly required only in the flipped and online approaches. Moreover, students who accessed online materials more often outperformed those who did not. Data on these measurements was not collected in our study and may have influenced our results, especially considering that both the C and E groups were consistently required to use online resources to complete homework and take assessments.

Thus, to further ascertain if the flipped learning pedagogy had any other tangible effects on minority students' achievement, we investigated data unique only to the 43 students in the E group. We analyzed the relationship (correlation) between the two variables x, the Chapter Notes grades (the independent or explanatory variable) and y, the Final Grade (the dependent or response variable). The Pearson coefficient r determines the strength of a linear relationship with a value of $-1 \leq r \leq 1$ (–1 signifying the strongest negative relationship with one variable going up as the other goes down, and +1 the strongest positive relationship with both variables going up or down). The correlation coefficient r^2 is the measure of the proportion of variation in the response variable that can be associated with the variation in the explanatory variable. The r-value of the correlation between the two variables was moderately positive (0.660), meaning that as Notes scores increased, a student's Final Grade in Elementary Statistics also increased. The r^2 for the relationship was 0.6915, suggesting that the variation between the two variables was closely associated. As r was sufficiently meaningful to justify creating a scatter plot, we made that graph and obtained the appropriate with regression model (see Graph 2).



Results related to Minorities' Math Anxiety and Self-Confidence

We now turn to our first research sub-question: Does the flipped learning pedagogy have an effect on minority students' math anxiety and self-confidence, both of which influence achievement? Students who have math anxiety display measurable behaviors that show they are feeling overwhelmed. One symptom identified by Sokolowski and Ansari (2017) is the lack of motivation to do math. This behavior is exhibited when students stop working on the self-directed portion of their coursework (homework); these students appear "frozen" in place.

Evidence revealed that 100% of students in both the C and E groups began the semester by completing every homework assignment. Then midway through the course (at midterm or 7 weeks), some students began to lose their impetus to work independently even though their attendance in class did not lag. (Attendance was not weighted towards the final grade, so they could have also started to skip class at this point, but absenteeism only occurred in one case out of the 92 students.) If a student was in the E group, they also continued to participate in the weekly group problem-solving Class Activities, even though they were apparently immobilized when working on their own. Frozen students seemed to be saving face by making the effort to look competent in front of their classmates, whereas in actuality, data revealed they had stopped doing individual work.

We collected data on the phenomenon of becoming “Frozen” by examining the point of the 14-week semester when students did not attempt to do any more homework – which gave the percentage of homework that was never attempted after that week of the semester. Students who skipped homework sporadically were not included in this compilation, because not completing homework on occasion cannot be attributed to the consistent lack the motivation common in math anxiety. Frozen behavior is differentiated by the act of “giving up” and feeling so overwhelmed that students stopped any attempt to work on their own.

The data was categorized demographically for both the C and E groups. Out of the total 92 students enrolled in the C and E groups, 30 self-identified as Black males (32.6%). The remaining 62 students (67.4%) were classified as “Other,” which included 50 Black females (54.4%) and 12 who self-identified as White, Hispanic, or Asian (13.0%) of any gender. The table below sorts the Black males out of the total number of so these results can be revisited later in this paper when that particular theme is discussed.

Table 4: Math Anxiety Expressed as Lack of Motivation to Do Homework

	Frequency of Frozen Students	Mean Week Frozen out of 14 Weeks = % of Work Never Attempted	Combined Frequency of Frozen Students
Control *n* = 49	Black Male (*n* = 15) $\frac{10}{15}$ = 66.7%	$\frac{11}{14}$ = 21.4%	$\frac{18}{49}$ = 36.7%
	Other (*n* = 34) $\frac{8}{34}$ = 23.5%	$\frac{12.6}{14}$ = 10.0%	
Experimental *n* = 43	Black Male (*n* = 15) $\frac{7}{15}$ = 46.7%	$\frac{10}{14}$ = 28.6%	$\frac{11}{43}$ = 25.6%
	Other (*n* = 28) $\frac{4}{28}$ = 14.3%	$\frac{12.5}{14}$ = 10.7%	

Several observations can be derived from the information in Table 4.

1. The results illustrate the high frequency of minority students who felt overwhelmed in Elementary Statistics before the end of the 14-week semester and therefore stopped working independently. Although the proportion of students who never undertook their end-of-the-semester assignments ranged from 14.3% to 66.7%. It must be noted that in both groups, the Other students who became Frozen were all Black females. Not a

single White, Hispanic, or Asian male or female exhibited Frozen behavior (although they represented 13.0% of the total students from the two samples).
2. Most students were not classified as Frozen. In 3 categories out of 4 (C-Other and both E), the majority of students (> 50%) persisted and never exhibited a lack of motivation.
3. Regardless of the pedagogy, more Black males were Frozen compared to their peers. In the Control, 43.2% more Black males stopped working before the end of the semester; in the Experimental, 34.4% more Black males were Frozen.
4. The mean percentage of homework assignments that Frozen students left undone was at least 10%, meaning that they stopped working for *weeks* before the semester was over and not just for a few days. Frozen Black males never attempted close to one-fourth (i.e. 21.4% and 28.6% respectively) of their independent assignments.
5. Black males gave up sooner than Other demographics in both the C and E groups. In the C group, they never attempted 11.4% more of their work than others; in the E group, they never attempted an additional 17.9% of their homework compared to peers.
6. The flipped learning pedagogy decreased the frequency of students who stopped doing their work in both demographics. For Black males, 20% fewer could be categorized as Frozen, and for Other, the in frequency of becoming Frozen was 9.2% less.
7. The overall frequency of any student being identified as Frozen was lowered by 11.1% when students participated in the flipped learning instructional model.

A second hallmark of math anxiety is being unable to recall concepts on a test when the student knew them before the exam, as evidenced by the requirement of earning a minimum 70% Pretest score (when a student displays this behavior, they call it "going blank"). Data on this conduct was obtained by examining the number of questions students left blank (unanswered) on the three tests in Elementary Statistics. As Table 5 illustrates, the model of instruction did make a difference in outcomes. C group students, who were taught via traditional lectures, left considerably more questions blank on every test compared to those learning statistics by the innovative flipped learning pedagogy. Overall, the C students left 11.9% of the questions on their tests blank, or 5.3% more questions unanswered than the E group. This result points to a supposition that the flipped learning components of Notes and the weekly problem-solving Class Activities enhanced student's self-confidence. E group students more often attempted a test problem rather than leaving it blank (which did not necessarily equate to the correct answer!)

Table 5: Number of Questions Left Blank (Unanswered) on Tests

	Test 1 (mean # Unanswered out of 42)	Test 2 (mean # Unanswered out of 29)	Test 3 (mean # Unanswered out of 36)	Total (mean # Unanswered out of 107)
Control	3.78	4.02	4.95	12.76 = 11.9%
Experimental	1.24	1.81	4.04	7.10 = 6.6%

Taken together, the results in Table 4 and 5 address our first research sub-question: Does the flipped learning pedagogy have an effect on minority students' math anxiety and self-confidence, both of which influence achievement? The evidence supports a positive change in disposition with regard to both traits. The innovative pedagogy decreased the proportion of students in all demographics who became Frozen (lack of motivation to do independent work) before the end of the 14-week semester. Furthermore, flipped learning gave students the self-confidence to attempt answering more test questions, rather than leaving them blank.

These results are relevant to our main research question: Does the flipped learning pedagogy have an effect on minority students' achievement in a college-level Elementary Statistics course? Although prior analysis showed there was no significant difference in the C and E groups' mean Final Grades in the course, the flipped learning approach appeared to have other measureable, positive effects in terms of math anxiety and self-confidence – and both of these attributes have an impact on achievement. As stated in the Literature Review section, math anxiety has a negative effect on achievement; thus, if flipped learning decreased an individual's level of anxiety, then that student's achievement most likely improved. In addition, as stated in the large meta-analysis done by Muhtadi and colleagues (2022), self-efficacy and the ability to achieve are positively correlated. Consequently, when the flipped learning pedagogy enhanced self-confidence, achievement would have increased simultaneously.

Our second research sub-question asked: Is there any identifiable difference in Black males' achievement when they participate in the flipped learning pedagogy? In the previous discussion of Table 4, statements 3 to 6 acknowledged the substantial gap between Black Males performance compared to others. The data revealed that more Black males were Frozen than any other demographic, and that they became Frozen earlier on than anyone else who also became Frozen. That evidence was indicative of a higher occurrence of math anxiety and a greater lack of self-confidence.

The final grades of students corroborate the gap in achievement for Black males. Table 6 contrasts the mean and median scores for the combined C and E groups' final grades:

Table 6: Combined Final Grades in Elementary Statistics

	Black Males	Other	Difference
Mean % (Out of 100%)	70.99	77.64	6.65%
Median % (Out of 100%)	70.52	78.36	7.84%

We can see that Black males earned grades that were about 7% lower than their peers who were taking the same course. This observation is consistent with findings stated earlier in the Literature Review regarding the math achievement gap between Black males and other groups. However, the combined grades also show that the curriculum provided in both pedagogical approaches (*PowerPoint* slides, Pretests, study sessions) resulted in moderate success for Black males. The mean and median of their final grades were both above 70%; Black males at our HBCU reached the benchmark of "passing" in Elementary Statistics.

So, is there evidence to show that *only* the flipped learning pedagogy had an effect on Black males' math anxiety and heightened lack of self-confidence? The strongest support for a positive effect in these two areas results from data that showed 20% fewer E group Black males were categorized as Frozen when compared to C group Black males (see Table 4). An innovation that leads to a decline in any behavior for one-fifth of the sample is meaningful.

Results Related to Non-STEM Majors Standard for Success

We asked a third research sub-question: Does the attitude that the below-average performance is not only sufficient, but also acceptable, in some disciplines affect student achievement, making it difficult to cultivate a diverse, competent population of professionals? To explore this question, data on the frequency of final grades for STEM and non-STEM majors is below.

Table 7: Frequency Distribution of Final Grades in Elementary Statistics					
STEM Majors *(Agribusiness Management, Biochemistry Biology, Business, Computer Science, Environmental Science, General Agriculture, Human Ecology)*			Non-STEM Majors (Exercice Science, Rehabilitation Psychology, Rehabilitation Services, Criminal Justice, Sports Management, Special Education)		
	Frequency	Distribution		Frequency	Distribution
Control *n* = 32	A: 6	6/32 =18.75%	Control *n* = 17	A: 1	1/17=5.88%
	B: 13	13/32=40.63%		B: 4	4/17=23.53%
	C :6	6/32 =18.75%		C: 9	9/17=52.94%
	D: 7	7/32=21.87%		D: 2	2/17=11.76%
	F : 0	0%		F: 1	1/17=5.88%
Experimental *n* = 27	A: 3	3/27=11.11%	Experimental *n* = 16	A : 0	0%
	B: 7	7/27=25.93%		B: 4	4/16=25%
	C: 10	10/27=37.04%		C: 7	7/16=43.75%
	D: 4	4/27=14.81%		D : 3	3/16=18.75%
	F: 3	3/27=11.11%		F: 2	2/16=12.5%
n = 59			*n* = 33		

As mentioned in the "Minority, Education, Mathematics and Statistics" section earlier in this paper, some students at HBCUs tend to struggle with Elementary Statistics because of a lack of emphasis on meeting sufficient math requirements prior to taking the course. This situation particularly applies to non-STEM majors who must pass only one low-level course similar to high school Algebra 1. Observations of the combined C and E groups' final grades presented in Table 7 call attention to some interesting facts:

1. STEM majors are required to pass College Algebra (rather than a lower-level Algebra 1 equivalent) before enrolling in Elementary Statistics. Therefore, whether they were in the C or E groups, they come into the course with greater competence in mathematics, as evidenced by their attaining the highest frequency of A's and B's: $\frac{29}{59} = 49.2\%$
2. Non-STEM majors (who take the low-level Algebra 1 equivalent) exhibit a considerably lower frequency of A's and B's: $\frac{9}{33} = 27.3\%$, indicative of their inadequate preparation.
3. STEM majors have a frequency of a Cs: $\frac{16}{59} = 27.1\%$, whereas non-STEM majors have a higher frequency of Cs: $\frac{16}{33} = 48.5\%$. The difference in the number of students willing to tolerate an average performance is suggestive of their perception about the standard that is deemed "acceptable" by their major department, as well as what they believe will be important in their eventual professional career.
4. As STEM majors do not allow grades lower than a C in Elementary Statistics to count towards graduation (judging the performance sub-standard), those low-achieving students with D's or F's, or $\frac{14}{59} = 23.7\%$ are required to retake the course.
5. Non-STEM majors earning a D in the course in both groups, or 11.76% and 18,75% of them respectively, will be allowed to use that below-average performance towards graduation, and none of them will retake Elementary Statistics. Only F's will repeat.

Points 3 and 5 are especially troubling because Elementary Stats is the foundational course that prepares today's students to interpret the plethora of data gathered in all human endeavors, and teaches them to make well thought-out decisions from that data. Certainly, one would presume the major fields listed as Non-STEM in Table 7 would want to cultivate a diverse, competent population of professionals who can think logically and solve problems inherent in their discipline. However, the evidence of final grades suggests that the non-STEM departments at our HBCU (and possibly others) have created a climate where statistics is thought of as less valuable than other subjects in a student's program of studies. At a minimum, it seems they have advanced the belief that statistics is too difficult for students in their majors because a substandard performance is thought acceptable. Such attitudes are counter-productive in closing achievement gaps, fostering math self-confidence, or creating opportunities to learn. When students arrive to class saying, "I'll be happy with a D in stats, that's all I need to graduate," they are predisposed to fulfilling the cycle of low achievement discussed previously.

CONCLUSIONS

We were interested in learning if an innovative instructional model called "flipped learning" would have an effect on minority students' achievement in a college-level Elementary Statistics course at an HBCU (Historically Black College and University).

Conclusions Related to Final Grades (Achievement)

The results of this study showed that when two dissimilar pedagogies were used in teaching (lecture and flipped learning), there was no significant difference (p = 0.2494) in the final grades for two comparable groups of students (the Control C and Experimental E) when other instructional resources were the same (e.g. the instructor, assignments, and assessments). Descriptive measures and a box plot underscored the similarity in final grades. This result might have occurred because in both groups, students had to earn a 70% or more on Pretests. These preparatory assignments, coupled with weekly Study Sessions, apparently did an excellent job of coaching students in the skills they needed to succeed in subsequent high-stakes assessments. As the Pretests, Tests, and the Final Exam accounted for a significant proportion (60%) of the student's final grade in the course regardless of the pedagogy, the impact of the flipped learning with its innovations of Notes and Class Activities appeared to have minimal influence on final grades. As discussed previously, other researchers (Burke & Fedorek, 2017) found a similar lack of flipped learning's impact on achievement in their studies.

However, evidence indicated that when the E group took pre-class Chapter Notes and participated in weekly group problem-solving Class Activities, it positively affected their overall grade. The Pearson correlation coefficient r = 0.660 and the correlation coefficient r^2 = 0.6915 allowed us to create a scatter plot and obtain a linear regression model. We were able to conclude that when students' scores on Notes increased, so did their final grade. Thus our investigation aligns with many previous studies that showed the flipped learning pedagogy does have some positive effect on achievement.

Effect on Math Anxiety and Self-Confidence

Our first research sub-question examined the flipped learning pedagogy's effect on minority students' math anxiety and self-confidence in Elementary Statistics. Evidence showed that the innovative pedagogy decreased math anxiety, as measured by a student's Frozen behavior (becoming so overwhelmed that they lost motivation to do math). We found that participating in the flipped learning instructional model lowered the frequency of being identified as Frozen by 11.1%. To determine the effect of the innovative instructional model on self-confidence, we investigated the number of questions students left blank (unanswered) on tests. Overall, the C (lecture) students left 11.9% of the questions on their tests blank, or 5.3% more questions unanswered than the E (flipped) group. This second result points to a supposition that the flipped learning components of Notes and the weekly problem-solving Class Activities enhanced student's self-confidence. E group students were at least willing to try a test problem rather than leaving it blank.

Effect on Black Males

Our second sub-question explored the effect of flipped learning on Black males. We discovered that Black males (regardless of pedagogical approach) tended to have a higher percentage of incomplete work compared to the overall population, as well as giving up on their studies sooner. However, flipped learning reduced the proportion of Black males exhibiting this behavior by 20%. Further evidence revealed that Black males' grades were approximately 7% lower than all other students' grades. However, Black males mean and median scores showed they were passing the course, regardless of pedagogy, so the preparatory work completed by everyone (Pretests and Study Sessions that matched subsequent assessments) were the key to success.

Conclusions about the Choice of College Major

In our final sub-question, we considered the attitude at some HBCUs that allows non-STEM majors to graduate with below-average performance in Elementary Statistics. An analysis of data collected on the grades of different majors showed that STEM majors had the highest frequency of A's and B's (49.2%), whereas Non-STEM majors had a greater percentage of C's (45.2%) and permitted D's to count for graduation. This evidence points to a climate where statistics is thought of as less valuable than other subjects in a students' program of studies, or the belief that statistics is too difficult for students in these majors. Such attitudes are counter-productive in closing achievement gaps, fostering math self-confidence, or creating opportunities to learn.

Final Decision with its Implications and Limitations

To make a decision about the worth of the innovative flipped learning pedagogy and its effect on minority students' achievement in college-level Elementary Statistics at our particular HBCU, we returned to the Decision Theory discussed earlier in this paper. According to Slovic, Fischhoff, and Lichtenstein (1977), after examining the normative and descriptive evidence from our study, we should "take the course of action that conforms most closely to the decision maker's beliefs and values." Thus, we are justified in making an educational decision sensitive to our value judgements about what counts as improvement with regard to the teaching and learning of our minority students. We acknowledge that the impact of flipped learning on final grades did not emerge as statistically significant when compared to a lecture model. However, in other quantifiable areas, flipped learning had a positive influence on issues that challenge the minority population. Analysis showed that math anxiety diminished and self-confidence was enhanced. Black males, in particular, showed meaningful improvement in negative behaviors like becoming Frozen. Furthermore, flipped learning encouraged the development of four of the "capacities" that allow students to flourish (as described in Brighouse et al, 2018).

Our HBCU helped with in the decision-making process by listening to instructor and student input, and then addressing key needs that arose from those discussions. A federal grant initially funded Teaching Assistants and secured a permanent classroom with circular tables and computers to support the flipped learning group environment. The grant also purchased a sufficient number of “loaner” calculators for use by economically disadvantaged students. After weighing our HBCU’s stated values and evidence found in this study, our Math Dept. decided that the flipped learning model should be the preferred pedagogy choice for teaching Elementary Statistics. In the next academic year, 75% of the available seats in the course are in sections that use this innovative instructional model. They also agreed to purchase additional calculators and fund Teaching Assistants for this course once the federal grant expires.

However, we recognize that the variables studied in this experiment may be relevant only to the minority populations at U.S. Historically Black Colleges and Universities (HBCU’s) or other institutions of higher education with similar student demographics, GPAs, and college entrance exam scores. Because other studies cited in this paper found both similar and conflicting results when they compared flipped learning to the traditional lecture approach at the college-level, any instructor considering the approach should first clearly outline their particular decision-making “value statements.” Any teaching and learning approach needs to be aligned with the expectations of the university, its culture, and its mission statement. In addition, adjusting curriculum from a traditional lecture format to the flipped learning pedagogy requires the teacher to make a considerable commitment in terms of time and energy, so there must be a clear alignment to both the university and instructor’s values/beliefs to warrant the sacrifice.

Finally, we believe the ultimate goal of any institution of higher learning is to cultivate a diverse and competent population who is able to make sound decisions in today’s world. Since Elementary Statistics offers the foundational coursework for developing the skills that are valued in decision-making, our HBCU is committed to preparing economically productive graduates who can make and act on well-informed and well thought-out judgments. Therefore, we also hope that other HBCUs and minority-serving 2-year and 4-year colleges reading this study will consider testing this innovative pedagogy on their campuses.

APPENDIX 1: EXAMPLE CHAPTER NOTES ASSIGNMENT (AKA NOTES)

<u>Chapter 3 Notes</u>

Notes 3a) On your own lined paper, due at the <u>beginning</u> of class on date stated on Assignment Calendar:

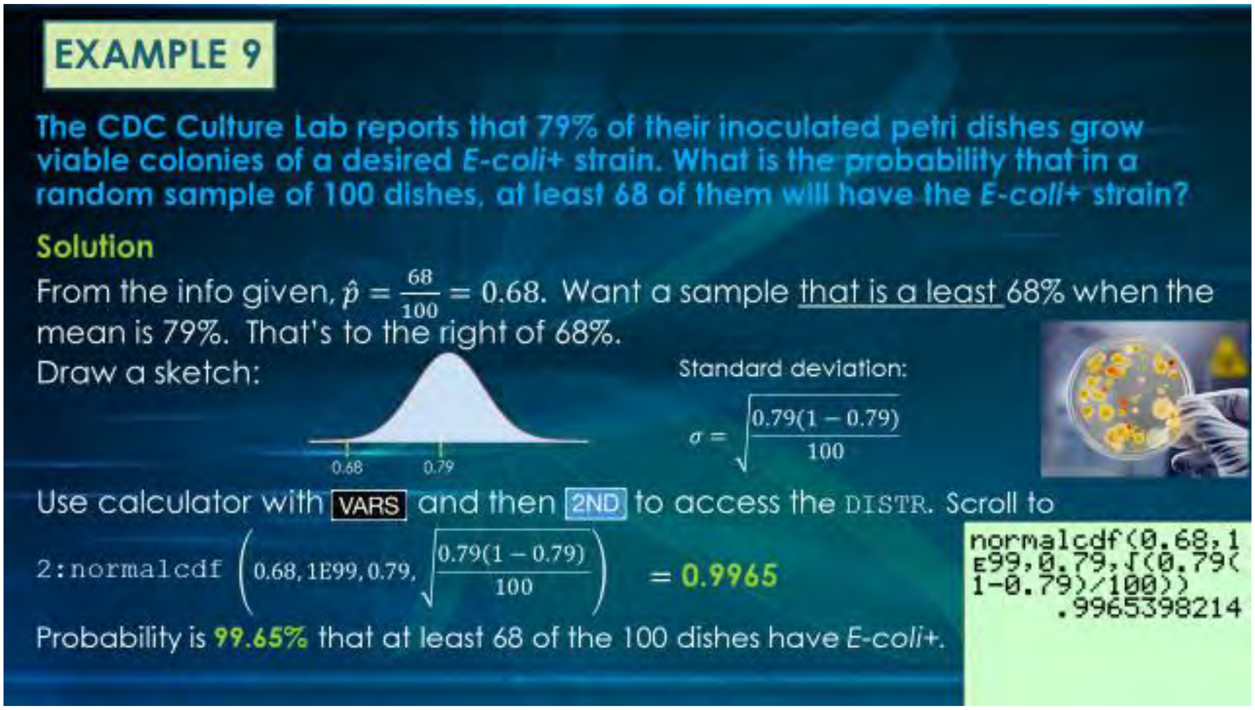
1) From Section 3.1: Copy the chart called "Summary of Measures of Center" on slide #24.
2) From Sec 3.2: Write the definitions of all terms on the two slides #26 - 27 on "Dispersions", and slide #32 "Coefficients of Variation"
3) Watch this video: https://m.youtube.com/watch?v=179ce7ZzFA8 (*8.25 min.*)
   Write a minimum of 5 sentences about what you learned.
4) From slide #34, along the <u>left side</u> of your paper, draw the three "Empirical Rule" bell-shape curves with percentages inside them. On <u>the right side next to each curve</u>, write the sentence describing the percentage (starts with "Approximately…")

Notes 3b) On your own lined paper, due at the <u>beginning</u> of class on date stated on Assignment Calendar:

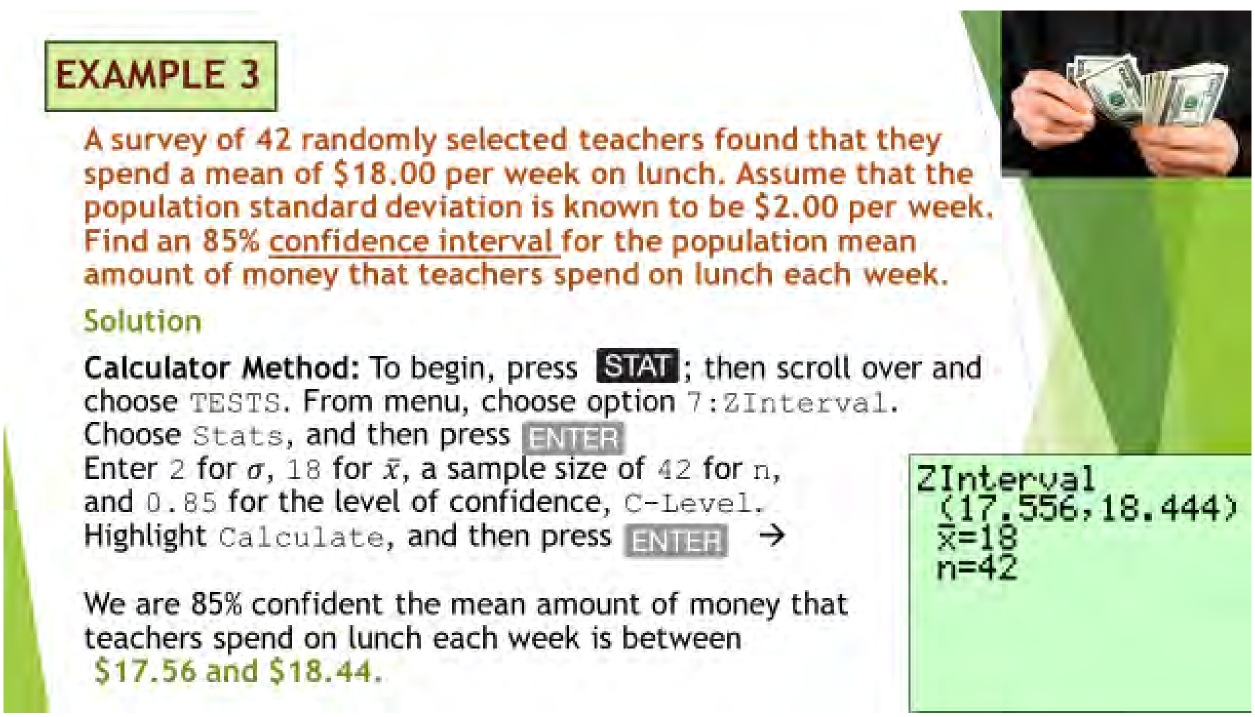
1) From Section 3.3: Write the definitions of all terms on the two slides #42 - 43 "Quartiles" <u>and</u> slide #48 "Standard Scores"
2) Watch this video: https://www.youtube.com/watch?app=desktop&v=fJBSgkbUuFA (*6.33 minutes*)
   Write a minimum of 5 sentences about what you learned.
3) An in-class *Group Activity* will be added to the grade on Notes 3b.

APPENDIX 2: EXAMPLE POWERPOINT SLIDES

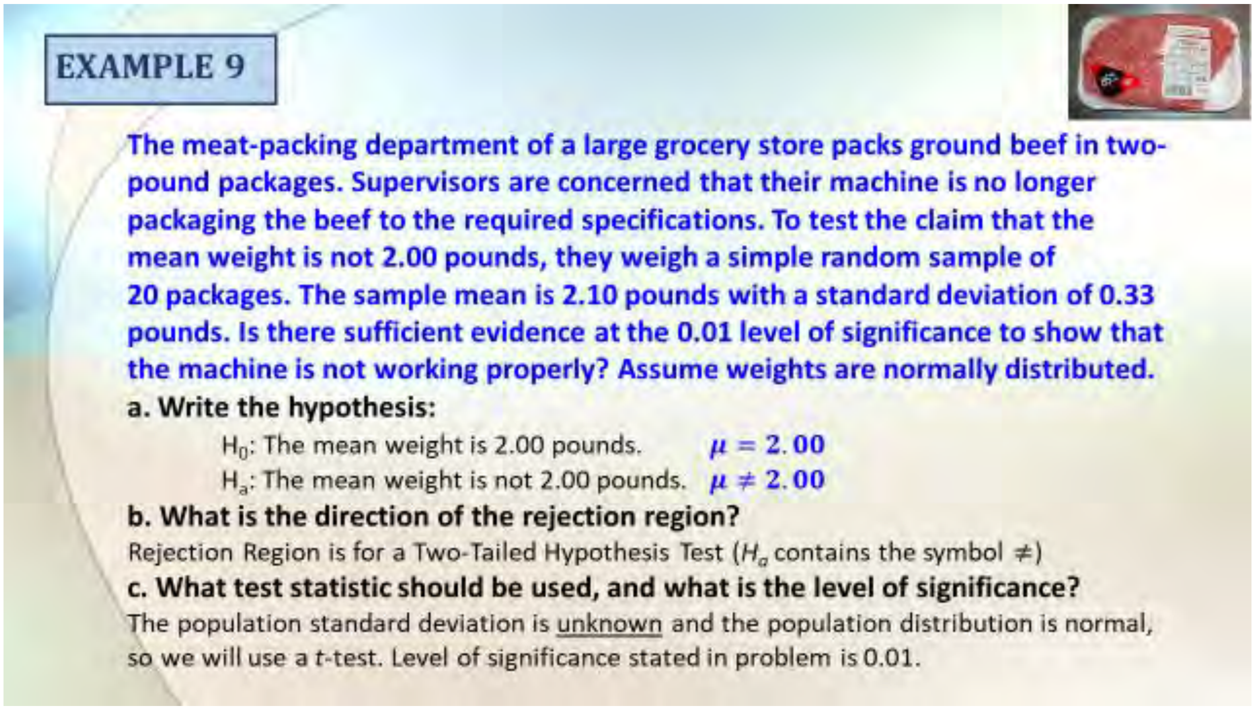
Example 1: Finding the Probability of a Proportion



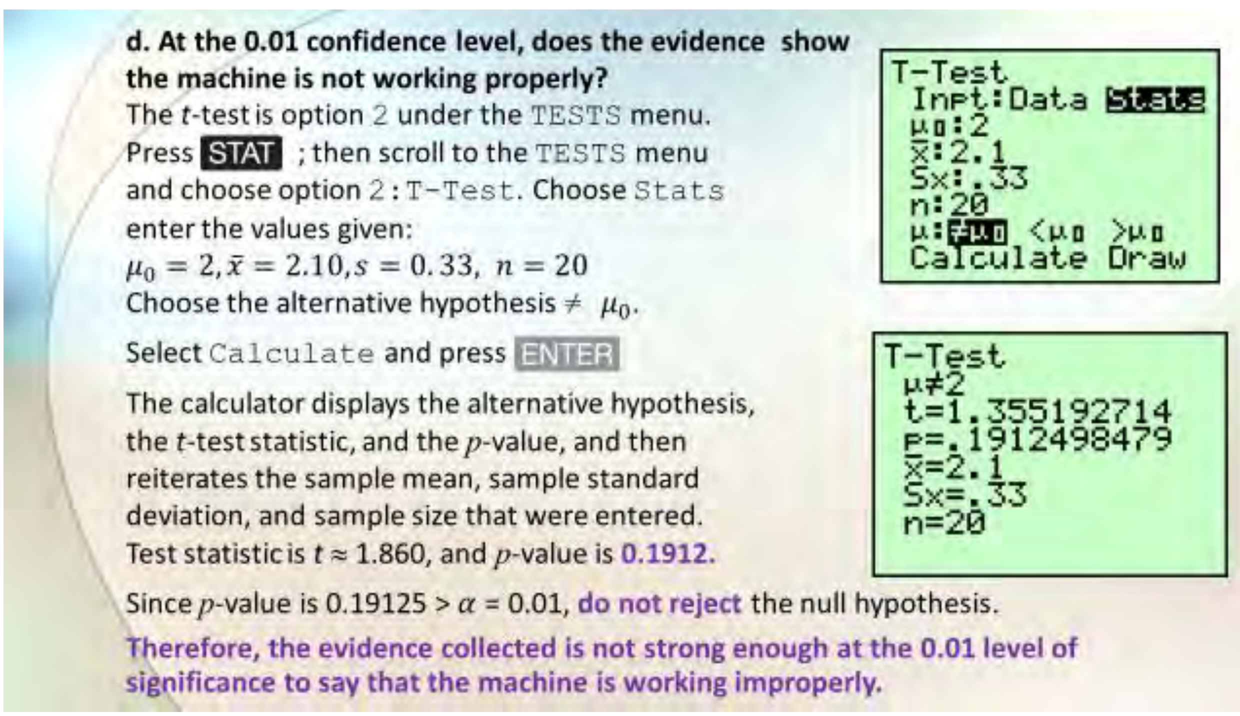
Example 2: Constructing a Confidence Interval for a Population Mean (σ known)



Example 3: Testing Hypotheses with t-statistic (Slide 1)



Example 3 Continued: Testing Hypotheses with t-statistic (Slide 2)



APPENDIX 3: EXAMPLE CLASS ACTIVITIES

Example 1: Consider the Source When You Read a Study

Class Activity 1b

Woman's World magazine has a circulation of over 1.3 million readers. Page 17 of the January 11, 2021, issue contained the article with the headline and excerpt below:

> **Headline:** Japanese breakthrough lowers dementia risk by 64%. STAY SHARP!"
> **Excerpt:** "Good news for chocolate lovers: Sip 2 mugs of hot cocoa daily and a study in The American Journal of Clinical Nutrition found you'll up your "good" cholesterol by 24%. Researchers credit cocoa compounds called flavonoids with the heart-smart effect. What's more, Appalachian State University scientists say folks who enjoy 25 grams of dark chocolate (about four squares) daily have better memories and multitasking ability than those who don't indulge."

The study *Woman's World* magazine used to write this article came from The American Journal of Clinical Nutrition and can be found at DOI: https://doi.org/10.3945/ajcn.114.092189 Volume 101, Issue 3, March 2015, Pages 538–548.

Your Task: Work with your group – but each person must turn in their own paper for points!

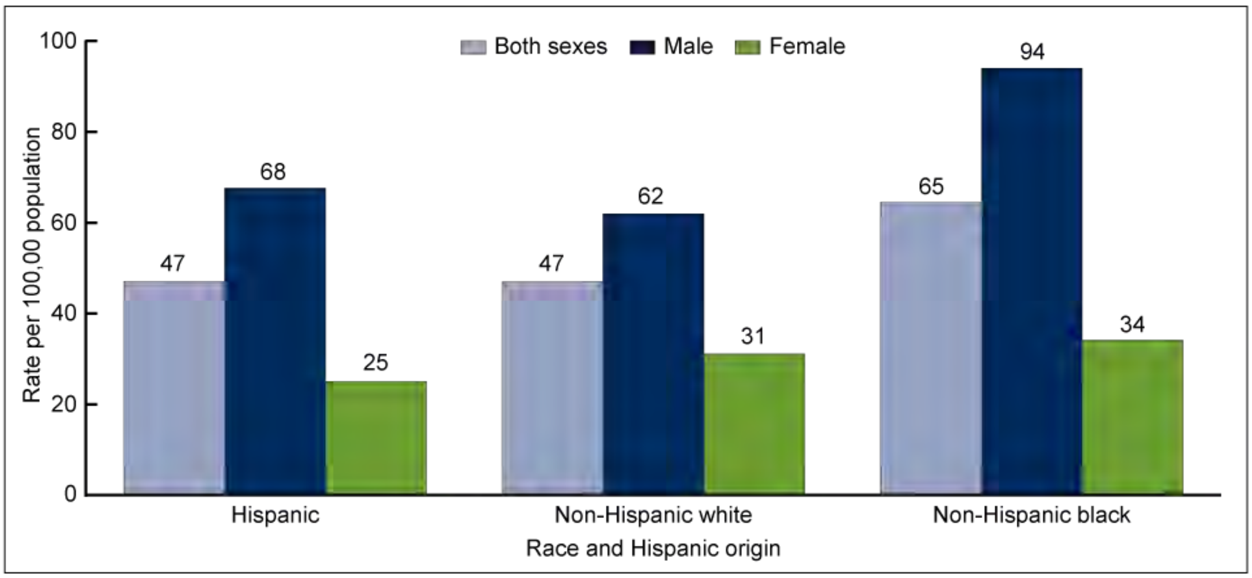
- Read the original study abstract (pull up the link on your phone or machine) - which includes background, objective, design, results, and conclusions.
- Describe (from the original study):
  - the type of study:
  - the population:
  - the sample:
  - list other important study details
- Is the headline and excerpt from *Woman's World* supported by what you see in the abstract of the original study they say it came from? Why or why not? What is your opinion of the ethics of this situation (for *Woman's World* readers)?

Example 2: Calculating Probabilities

Class Activity 4b

From 1999 to 2006, an average of 16,375 teenagers 12-19 years died each year. The overall risk of dying for teenagers (average annual death rate) is 49.5 deaths per 100,000 population. Accidents (unintentional injuries), homicide, suicide, cancer, and heart disease make up the five leading causes of death for teenagers. Motor vehicle fatality is the leading cause of accident death among teenagers, representing over one-third of all deaths to teenagers. Homicide is the leading cause of death for non-Hispanic black male teenagers, with more than two of every five deaths due to homicide. [*From CDC* (*Center for Disease Control*) https://www.cdc.gov/nchs/products/databriefs]

Figure 3. Death rates for teenagers 12–19 years, by race, Hispanic origin, and sex: United States, 1999–2006



SOURCE: National Vital Statistics System, Mortality.

1) Create a table for the data for males and females only (*do not use the "Both sexes" bar*) of the rates per 100,000 below (similar to what you see on slide #13 of Sec 4.2, Example 3).

Consider these Events. *Then turn the paper over to answer questions about them.*

A = teen selected is male
B = teen selected is Non-Hispanic Black
C = teen selected is Non-Hispanic Black and female

2) List the names of events that are mutually exclusive and explain why in words:

   List the names of events that are **not** mutually exclusive and explain why in words:

3) Match the choices for probabilities of two events by drawing a line between them:

Dependent Events, Not exclusive	Use Addition Formula with subtraction
Dependent Events, Mutually exclusive	Use a Multiplication Formula
Independent Events, with replacement	Use Addition Formula with no subtraction
Independent Events, without replacement	Use a Multiplication Formula with division

Write the two Addition Formulas here:

Not exclusive:

Mutually exclusive:

Write the two of the Multiplication Formulas here:

With replacement:

One of the formulas for without replacement:

4) Find the probability (of the rate per 100,000) that a teen selected at random is either male or Non-Hispanic Black. Use the formula and show your work. *Round to nearest whole percent.*

5) Find the probability (of the rate per 100,000) that a teen selected at random is male and Non-Hispanic Black Use the formula and show your work. *Round to nearest whole percent.*

6) Find the probability (of the rate per 100,000) that a teen selected at random is Non-Hispanic Black and female. Use the formula and show your work. *Round to nearest whole percent.*

7) Explain in your own words: How did you decide which formula to use when asked to find a probability?

Example 3: Writing and Testing Hypotheses

Class Activity 10a

The famous Titanic left Southampton, England towards New York on Wednesday, April 10, 1912. Many of the rich who boarded there thought the Titanic could not sink. The ship made brief stops in Cherbourg, France, and Queenstown, Ireland to take on third class passengers, and then set its course across the Atlantic. On the evening of April 14, it struck an iceberg, and in less than two hours, by 2:15 AM, it sunk. Only 710 survived among 2201 passengers and crew. A question of interest is whether males and females had the usual chance of securing a seat on a lifeboat (it appears the Titanic ratio was 1.4 women per 1 man). A simple random sample of 128 shipwrecks shows the mean population survival rate is 1.5 women per 1 man, with a standard population deviation of 0.8. With 90% confidence, was the Titanic different?

(a) State the hypothesis in words and using the μ relationship:

(b) The direction of the Rejection Region is for a _______ Tailed Hypothesis Test because the H_a contains the ___________ sign.

(c) What measures (criteria) are stated that help you decide which test statistic to use?

What is that test statistic? _________ What is the level of significance? ____________

(d) Will you reject H_0 at the level α = 0.10 if mean survival ratio is more or less than 1.5? To answer this question, use your calculator and try both of the methods of hypothesis testing.

- **First, determine the answer using Rejection Regions**.

Use Table 10.1: Critical *z*-Values for Rejection Regions to find these values:

The critical z-value for α = ____________ from the Table is z_a: _________________

Draw a sketch with shaded portion, showing z_a on the x-axis and the Rejection Regions with arrows:

You will reject H_0 if ____

(Turn this paper over to continue working)

State what you entered into your calculator:

Answer you obtained for z (*to three decimal places*): ____________
Looking at your sketch, where would you put this answer?

In what Rejection Region do these values fall?

So, using Rejection Regions, would you reject **OR** fail to reject the H_0?

- **Next, determine the answer using a p-value.**

State what you entered into your calculator, if it was different. (If no difference, state that fact.)

Answer you obtained for p (*to four decimal places*): ________________

What is the value for α? ________________

Circle the relationship that enables you to make a decision in this particular situation:

If $p \le \alpha$ reject H_0 **OR** $p > \alpha$ do not reject the H_0

Write the p and α values using the correct inequality from the step above: ________________
So, using the p-value, would you reject **OR** fail to reject the H_0? Why?

Based on both of these hypotheses testing methods, **use the level of confidence to explain to the general public whether the evidence supports OR does not support the claim** that males and females had the usual survival rate at the shipwreck of the Titanic. (Do NOT mention hypothesis rejection! The average person does not understand that terminology.)

REFERENCES

Albert, M. & Beatty, B.J. (2014). Flipping the classroom applications to curriculum redesign for an introduction to management course: Impact on grades. *The Journal of Education for Business*, *89*(8), 419-424.

Beilock, S. L., & Willingham, D. T. (2014). Math anxiety: Can teachers help students reduce it? *American Educator*, *38*(2), 28–33. https://eric.ed.gov/?id=EJ1043398

Bredow, C. A., Roehling, P. V., Knorp, A. J., & Sweet, A. M. (2021). To flip or not to flip? A meta-analysis of the efficacy of flipped learning in higher education. *Review of Educational Research*, *91*(6), 878–918. https://doi.org/10.3102/00346543211019122

Brighouse, H., Ladd, H., Loeb, S., & Swift, A. (2018). *Educational Goods* ([edition unavailable]). The University of Chicago Press. Retrieved from https://www.perlego.com/book/1852103/educational-goods-values-evidence-and-decisionmaking-pdf

Burke, A. & Fedorek, B. (2017). Does "flipping" promote engagement? A comparison of a traditional, online and flipped class. *Active Learning in Higher Education, 18*(1), 11-24.

Chapelow, J. (2023). Statistic in math: Definition, types, and importance. *Investopedia*. https://www.investopedia.com/terms/s/statistics

Chen, C., Huang, N., Norman, H., & Hwang, G. (2022). Findings and implications of flipped science learning research: A Review of Journal Publications. *Interactive Learning Environments*, *30*(5), 949-996.

Decision Sciences Institute (DSI). (2023). Retrieved from https://decisionsciences.org/

Doo, M. Y. (2022). Understanding flipped learners' perceptions, perceived usefulness, registration intention, and learning engagement. *Contemporary Educational Technology*, *14*(1).

Fadol, Y., Aldamen, H., & Saadullah, S. (2018). A comparative analysis of flipped, online and traditional teaching: A case of female Middle Eastern management students. *The International Journal of Management Education, 16*(2), 266-280.

Jackson, L., Ford, J., Randolph, C., Schleiden, C., Harris-McKoy, D., & McWey, L. (2021). School climate as a link between high school black males' math identity and outcomes. *Education and Urban Society*, *53*(4), 469-487.

Johnson, D., Johnson, R., & Stanne, M. (2000). *Cooperative learning methods: A meta-analysis.* Minneapolis, Minn.: University of Minnesota. July 2005: http://www.clcrc.com/pages/cl-methods.html.2000

Jung, H., Park, S. W., Kim, H. S., & Park, J. (2022). The effects of the regulated learning-supported flipped classroom on student performance. *Journal of Computing in Higher Education*, *34*(1), 132-153.

Khasawneh, E., Gosling, C., & Williams, B. (2021). What impact does maths anxiety have on university students? *BMC Psychology, 9*(37). (2021). https://doi.org/10.1186/s40359-021-00537-2

Luttenberger S., Wimmer S., & Paechter M. (2018). Spotlight on math anxiety. *Psychology Research and Behavior Management*, *8*(11), 311-322. doi: 10.2147/PRBM.S141421. PMID: 30123014; PMCID: PMC6087017.

Martin, D. B. (2000). Mathematics success and failure among African-American youth: *The roles of sociohistorical context, community forces, school influence, and individual agency*. Mahwah, NJ: Erlbaum.

Mills, B. J. (2012). IDEA Paper No. 53: Active learning strategies in face-to-face courses. Manhattan, KS: The IDEA Center. http://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/ PaperIDEA_53.pdf

Muhtadi, A., Assagaf, G., & Hukom, J. (2022). Self-Efficacy and students' mathematics learning ability in Indonesia: A meta-analysis study. *International Journal of Instruction*, *15*(3), 1131-1146.

Mulat, T., & Arcavi, A. (2009). Success in mathematics within a challenged minority: The case of the students of Ethiopian origin in Israel (SEO). *Educational Studies in Mathematics, 72*(1):77–92.

Murrell, P. C. (1995). In search of responsive teaching for African-American Males: An investigation of students' experiences of middle school mathematics curriculum. *Journal of Negro Education, 63*(4), 556–569.

Pringle, N. M., Workman, J. L., & Arrastia-Chisholm, M. C. (2022). Making connections to address mathematics anxiety: A case study of the instructional triangle and remedial college instructors. *Georgia Educational Researcher*, *19*(1), 1-21, Win 2022.

Reeves R. V., & Halikias D. (2017). *Race gaps in SAT scores highlight inequality and hinder upward mobility*. https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility/

Ridlon, C. L. (2009). Learning mathematics via a problem-centered approach: A two-year study. *Mathematical Thinking And Learning, 11*(44), 188-225.

Ridlon, C.L. (2013). Can active instructional strategies help college-age minority students succeed in math? *Annual Meeting of the Modern Technology and Management Institute*; September 20-21, 2013: Virginia Beach, VA.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, *28*(1), 1-39.

Sokolowski, H., & Ansari, D. (2017). Who is afraid of math? What is math anxiety? And what can you do about it? *Frontiers for Young Minds*, *5*(57).

Swigart, T. E., & Murrell, P. H. (2001). Factors influencing estimates of gains made among African-American and Caucasian Community College students. *Community College Journal of Research and Practice, 25*, 297-312.

Toldson, I. A., & McGee, T. (2014). What the ACT and SAT mean for black students' higher education prospects. *The Journal of Negro Education*, *83*(1), 1–4.

DECISION SCIENCES INSTITUTE

Forecasting Power Use via an Adaptive Neural Fuzzy Inference System – Primed by the Classification and Regression Tree Method

Justin L. Shoger
Department of Systems Science and Industrial Engineering, Binghamton University, State University of New York, Binghamton, NY 13902-6000, USA
Email: jshoger1@binghamton.edu

ABSTRACT

People need to make decisions in their everyday lives. Increasing amounts of data are available, characterizable as the glare of information: information requested vs. needed; and information provided offering greater freedom of choice. Increasing the number of choices does not make decision making easier. The paper's focus is on the usefulness of methods towards understanding and forecasting power use. With this prediction, one may improve their strategies and subsequent decisions in many areas such as financial budgeting by operating with less margin, lower their carbon footprint by reducing consumption, or balancing electrical power use with other means.

KEYWORDS: Adaptive neuro-fuzzy, artificial neural network, fuzzy inference system, classification and regression tree, decision making aid, resource allocation

INTRODUCTION

The ability for people to be able to forecast costs helps them better prepare for more effective and efficient allocation of resources. There are many options available to families in attempting to budget their monthly resources to cover costs of housing, sustenance, and clothing before deciding what the disposable income may be for recreation, hobbies, etc. Being able to manage expectations for power usage in order to heat and cool one's residence aids those families in better use distribution of their finite income. Previous application of a neuro and fuzzy approach to forecasting power use has been done by combining parallel processes of an artificial neural network (ANN) and a first order Takagi-Sugeno fuzzy inference system (FIS) to help make power consumption projections by the author in prior, unpublished research. In this study, the focus is on applying a neuro-fuzzy approach through an adaptive neuro-fuzzy inference systems (ANFIS) in developing a longitudinal model for electrical use of a private residence. Additionally, this study looks into the application of the classification and regression tree (CART) method for determining the entry membership function parameters for ANFIS processing. As well as producing dependable estimates for power consumption, the ANFIS approach allows the user to better understand the inner working of the calculations processes through to the outputs in contrast to the standard application of ANNs and their 'black-box' workings getting to the results.

LITERATURE REVIEW

Applying a multitude of approaches to aid in human decision making has included traditional analytic methods e.g. multivariate analysis, differential and partial differential equations, etc., and has broadened to include soft computing methods such as fuzzy logic, genetic algorithms, and others, meeting need for a better way to deal with ill-defined and uncertain situations and environments. (Zadeh, 1973 & 1994, Jang SM, et.al. 1997). The review of this work shows a progress of development from a focus on traditional computational methods to use of fuzzy models and approaches to human decision making such as fuzzy logic, fuzzy inference systems

(FIS), generalized neural networks, adaptive-network-based fuzzy inference systems (ANFIS), and other applications like classification and regression trees (CART).

Research into and application of fuzzy logic, fuzzy systems, and soft computing techniques has increased across an ever-widening spectrum of business and industry areas in the recent past. (Kar, et.al, 2014) Several additional approaches to CART have been investigated into developing training methods and parameter preparations to improve ANFIS computations such as Ant Bee Colony, Cat Swarm Optimization, Firefly Algorithm, Genetic Algorithm, and Particle Swarm Optimization. (Raja & Pahat, 2016) Application areas have and continue to include: student modeling systems, medical systems, traffic control, economic systems, image processing and feature extractions, forecasting and prediction, manufacturing and system modeling, social sciences, as well as research into further developing fuzzy calculation methods (Kar, et.al., 2014) The CART fed ANFIS hybrid offers a computationally lightweight and user understandable way of engaging in forecasting and / or prediction efforts that can be applied to utilities production, financial budgeting, and medical diagnosis. Specific CART-ANFIS applications include car following behavior in traffic control research (Moghadam, et.al, 2016); computer intrusion and attack detection and alarm (Pinem & Setiawan, 2015); industrial process fault diagnosis (Yang, et.al, 2009); and, as a non-parametric means to develop the ANFIS rules in occupational injury analysis. (Fragiadakis, et.al, 2014)

This section will present a synthesis of the research across the topics as it pertains to this paper's investigative approach. Table 1 provides a summary.

Table 1: Summary of Literature Review

YEAR	REFERENCES	JOURNAL
1973	Zadeh - Outline of a new approach to the analysis of complex systems and decision processes.	IEEE Transactions on Systems, Man, and Cybernetics
1991	Jang - Fuzzy modeling using generalized neural networks and kalman filter algorithm.	AAAI
1993	Jang - Input selection for ANFIS learning	IEEE Transactions on Systems, Man, and Cybernetics
1994	Jang - Structure determination in fuzzy modeling: a fuzzy CART approach	IEEE Conference on Fuzzy Systems
1994	Zadeh - Soft computing and fuzzy logic	IEEE Explore
1997	Jang, Sun, Mizutani - Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence	Reference Book
1998	Breiman – Classification and Regression Trees	Reference Book
2009	Yang, B. S., Oh, M. S., & Tan, A. C. C. Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference	Expert Systems with Applications
2014	Kar, S., Das, S., & Ghosh, P. K. Applications of neuro fuzzy systems: A brief review and future outline	Applied Soft Computing
2014	Fragiadakis, N. G., Tsoukalas, V. D., & Papazoglou, V. J. An adaptive neuro-fuzzy inference system	Safety Science

	(anfis) model for assessing occupational risk in the shipbuilding industry.	
2015	Pinem, A. F. A., & Setiawan, E. B. Implementation of classification and regression tree (CART) and fuzzy logic algorithm for intrusion detection system.	IEEE - 2015 3rd International Conference on Information and Communication Technology (ICoICT)
2016	Poor Arab Moghadam, M., Pahlavani, P., & Naseralavi, S. Prediction of car following behavior based on the instantaneous reaction time using an ANFIS-CART based model.	International Journal of Transportation Engineering
2016	Raja, P., & Pahat, B. A review of training methods of ANFIS for applications in business and economics	International Journal of u-and e-Service, Science and Technology

Fuzzy Logic and Soft Computing

As decision analysis applications grew and the need to incorporate human expertise in dealing with real-world issues became more sought after, ways to bridge the gaps between traditional analytical methods and a human centered approach were developed. These processes enabled inclusion of hard to quantify human reasoning methods and make that information useable in decision making and choice analysis. (Zadeh, 1973 & 1994) These new approaches looked to solve certain issues: converting human expertise to inputs for modeling in fuzzy reasoning models, i.e. fuzzy inference systems and the rules by which they compute; and, identifying the membership functions. Based on development of FIS with Takagi, Kang and Suegeno (TSK) towards real-world inference applications for systems' prediction and control (Jang, 1991 & 1993). Based on this groundwork, the discipline was advanced towards modeling systems with numerous inputs, non-normalized data and prioritize inputs, and extreme complexity which veiled both the systems' behavior and the modeling and analysis. The development of ANFIS combined neural networking with FIS fuzzy rules, hybridizing the methodologies for easier to use computation, faster learning, and quick, accurate results (Jang, 1991, 1993, 1994, & 1997)

Adaptive Network-Based Fuzzy Inference System

Focusing on methods to emulate human reasoning and how the systems developed from first ideas through to human-machine interactions, the approach synergized the calculus of fuzzy rules in interpreting human observations, and applying gradient descent and least squares errors methodologies to hone ANFIS performance and producing high machine intelligence in the process. The ANFIS model also sought to address application of the hybridized neural networks and fuzzy logic methods to identify systems' behaviors in predicting future outputs, explaining relationships and associated interactions across the inputs and outputs, and human decision-making modeled controller functionality. The results were a transparent computational model, incorporating human knowledge and inference supporting decision-making. The ANFIS model is based on a 5 layer architecture from data input through membership functions, into a modified neural network and relationship calculations made via fuzzy inference rules to an output. Through this work two problem areas emerged, transfer of human experience into the model, refining the parameters, and establishing input membership functions and the fuzzy if/then rules, identifying the structure. (Jang, 1991, 1993, 1994, & Jang, et.al. 1997).

Parameter optimization was achieved via backpropagation of the fuzzy rules output. This optimization process tuned the model's functions and behaviours and for pattern learning. This method was known to be slow and likely to get trapped in a local minima; and, including evaluation via least squares error provided a measure to gauge performance. This combination sped up training and results (Jang, 1993). Additionally, ANFIS provided some relief from depending on experts to assign the initial fuzzy rules, as the model curates them through training. (Jang, et.al., 1997)

Identifying the data structure for the model was the next challenge addressed. The consequent or fuzzy if/then rules require defined membership functions as the way to calculate from the linguistic variables and provide a mathematical output. Expert elicited membership functions set them for FIS models. The goal was to develop a means to reduce that dependency, and if possible improve speed and performance through training reduction. The CART method was applied to ANFIS to meet that requirement (Jang, 1993, Jang, et.al. 1997). CART methodologies were advanced by Breiman in the 1980's from Sonoquist's original 1960's work (Breiman, 2017). This work defined classifier and partitioning to uncover predictive structures in problem's data, and as a means to understand the variables and their interactions. CART applications provide insight into complexity of the data set with high dimensionality, mixed data types and non-standard data types. Classification helps expose relationships amongst variables as they change across the measurement space, and mitigates "the curse of dimensionality", per Bellman, 1961 (Breiman, 1998).

CART helps quickly determine rough and accurate structure of FIS enabling quick membership function learning for parameters and without a normalization layer (or need of human inputs). The number of fuzzy rules goes up exponentially with number of inputs, and choosing inputs with the most predictive power increases efficiency and accuracy. Classification and regression tree non-parametric method uses binary trees for error estimation via an impurity function. Combining ANFIS with CART compounds approaches to solving real-world problems in understanding pattern learning, model high non-linear functions, identify non-linear behaviours, and predict chaotic time series (Jang, 1994, Jang, et.al., 1997, Breiman, 1998)

APPLICATION BACKGROUND

This is a longitudinal study of a single family home. Several independent variables were available to illustrate the situation – year, month, monthly average temperature, pressure, and wind speed, and monthly cumulative rainfall over the course of nearly 20 years. Two dependent variables were identified and data collected, cumulative monthly power use and monthly power cost as billed by the local electric company. The independent variables chosen were monthly average temperature in degrees Fahrenheit (hereafter referred to as temperature) and monthly cumulative rainfall in inches (hereafter referred to as rainfall), being continuous throughout the study period. The dependent variable selected was monthly power use (hereafter referred to as power); power cost was not used in order to isolate the power prediction from cost inflation over the 20 year study. This research did not attempt to generalize this methodological application to any other case within the local geography or nationally. Other features of the residents have not been compared to normalized residential occupants which further constrains the general applicability of the predictions. However, the methodologies used in this investigation are definitely applicable to a broad range of prediction opportunities.

Power used at the studied residence, is the focus in applying the CART and ANFIS methods to understand and attempt to project power use based on the independent input variables, temperature and rainfall. This paper leverages the applications of artificial neural networks and fuzzy inference systems as they relate to the development and processes of ANFIS, and a more detailed description on the structure and methodology of ANFIS, and the purpose and applicability of CART as a precursor to ANFIS calculations. The following section presents the background on the context of the application and why it is a challenging problem. The approach and the specifics of the applications of the aforementioned methods to the investigation of the problem set data are covered next. The specifics of the methodologies applied to the data complete the investigation; and, the results, and conclusions, insights, and future study opportunities, complete the paper.

Resource Allocation Decision Making

Being able to make resource allocation decisions is important to everyone. With the ever increasing amount of information that is available for people to seek, and even more so, presented to them regardless of their desires or queries, people can be supported with tools to help them work through that deluge of information in a more systematic and coherent manner. Many available tools, such as machine learning and deep learning, are 'black box' models, opaque to the operator and/or the consumer of the analysis. That is, the inner workings are not well understood, or easily modified with clarity of subsequent changes by untrained users. To maintain awareness and progress despite excessive influxes of data, people need help in making projections of what the future may hold, and the means that the chosen aids absorb and process data with subsequent predictions. Understanding the algorithms, and the processes, the operator can make better choices within the context(s) the environment presents, as no amount of computational power can provide all the possible parameters for one's daily life.

Hybrid Algorithm – Adaptive Neuro-Fuzzy Inference System

ANFIS is a hybrid algorithm in that is combines gradient descent methods with in ANNs with least squares error methods for refining the consequents, fuzzy if/then rules. ANNs aid in clarifying the structure of the data available through the stochastic gradient descent method have been the focus of much attention in machine learning. These algorithms applied in conjunction with fuzzy logic, specifically Takagi-Sugeno fuzzy inference system, have been developed to map from independent variables' inputs to the dependent variable's output in optimizing the system's parameters.

Hybrid Algorithm – Classification and Regression Trees

The CART method is a supervised learning method allowing the resultant partitions of the data space, thus providing the overall initial data structure used as the membership functions within the ANFIS architecture. CART is a non-parametric method that features conceptual simplicity and computational efficiency. The current study used the regression tree methodology, specifically, for partitioning the continuous independent data, average monthly temperature and cumulative monthly rainfall and identify relevant ANFIS inputs. The goal of the regression tree is to split the data, iteratively in order to separate out, or eliminate inherent error in that section of the data space. The lower the residual error in a partitioned area, the more accurate the eventual predicted value will be. The CART process has many positive attributes, among them transparency of the process and results, the similitude to human decision making, and the graphical representation. These attributes can been seen in the calculation of the error values on either side of each split as shown in Fig. 1.

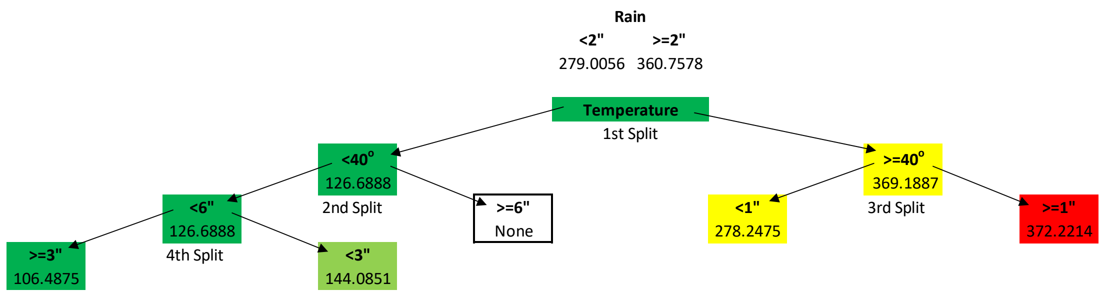


Figure 1. CART Regression Tree Reducing Estimation Error Branches and Leaves.

<u>Approach Background - CART</u>

The application of CART to ANFIS aids in structure identification of the initial architecture before parameter tuning by selecting relevant input variables. Specifically, using regression tree analysis in this study, the tree partitioning relieves the curse of dimensionality in that it narrows down the overall data space to well defined (mutually exclusive), applicable inputs. This has the added benefit of increasing the overall training efficiency of the combined CART and ANFIS applications. The main output of the using the CART method is determining the membership function parameters, α, β, and γ, making the process easier to follow by the implementer.
This study used the least squared error (LSE) method and the root mean squared error (RMSE) as the metric for prediction accuracy. The power output estimate was calculated as the mean across all the data points. Then the difference is calculated between the predicted and the actual output values, and then squares that difference, removing the resultant's valance, therefore providing a more accurate metric. The squared errors for all the data pairs are aggregated, divided by the number of data observations, arriving at a mean for the error value. This value has the square root taken, producing the RMSE value as the measure of prediction accuracy, using equation 1.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (1)$$

This bifurcation is repeated at intervals across the data set until a split produces the lowest inherent error on one side. Then, the focus switches over to the other data set, splitting that data space repeatedly in search of the lowest error value within that data set. When the lowest RMSE has been reached within each data set, the two are compared. Whichever independent data set has the lower RMSE is the root of the CART tree. From there the tree is grown via branches along the subsequent splits, ever reducing the RMSE of that data category. The splitting iterates between the two independent variables for a set number of levels, or until error is completely isolated, either of which can be the terminal leaves.

Once the terminal leaves have been established, the membership functions for the ANFIS portion of the hybrid algorithm were developed. Establishing the membership functions via the generalized Gaussian bell curve, see equation 2, is the preferred fuzzy inference system method. In this equation i and l are the independent variable data pairs across the entire data set; and α, β, and γ are the membership function parameters. The functions' shapes are dependent upon the data spread and number of partitions.

$$a_1[i,l] = \frac{1}{1 + \left| \frac{a_0[i] - \gamma[i,l]}{\alpha[i,l]} \right|^{2\beta[l,l]}} \quad \text{for } i = 1, \ldots, m,\ l = 1, \ldots, k[i]. \tag{2}$$

These characteristics contribute to the overall transparency of the CART method. In combining CART with ANFIS, the performance of standalone CART predictions is improved, as well as resilience mitigating data changes. There are additional supporting methods to improve lone CART analysis, but not used in this study.

Approach Background - ANFIS

The approach used for this study is different than standard ANFIS implementation by using both CART and ANFIS. The method chosen leverages several aspects including a feedforward ANN learning of a regular ANFIS weighted by the consequent rules. And, instead of relying on subject matter expert linguistic definitions for the membership functions, the CART methodology is used to initialize the membership functions, and assign membership grades to the independent variables. The CART process of partitioning of the data in order to reduce the estimated error values enhances the ANFIS hybrid algorithm accuracy and precision.
The structure of the CART and ANFIS approach, as shown in Fig. 2, was altered from the general, 5 layer ANFIS application. The architecture as implemented was 4 layers beyond the base, data layer. The 0 (base) layer is the independent variable input data. The first layer is the assignment of membership grades to the input data per the CART process. The next, second, layer is applying the membership graded data to the antecedent rules. The third layer is the application of the consequent coefficient ('c') values to the rules' outputs. The final and fourth layer aggregates the results of the consequents' weights for each input data point, resulting in a predicted output.

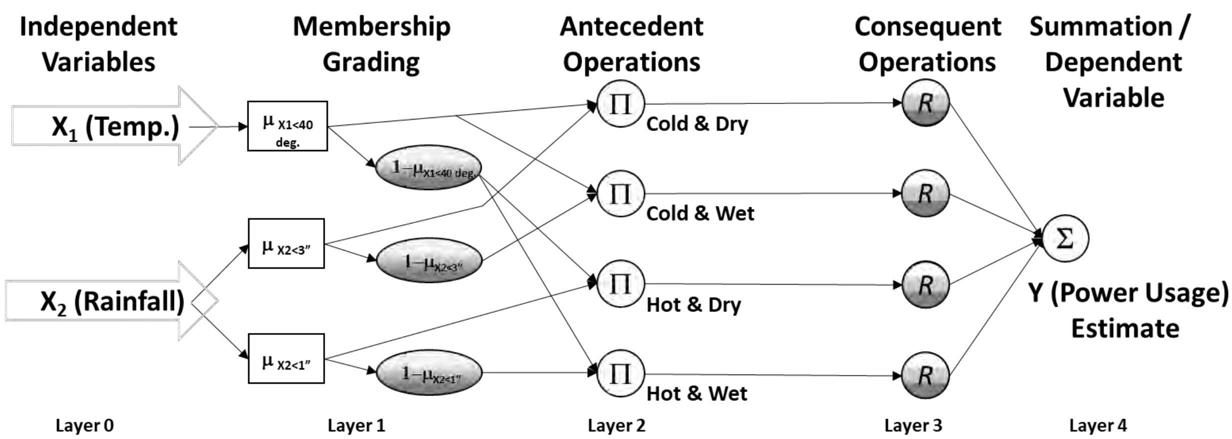


Figure 2. Implemented CART-ANFIS Structure.

As mentioned above the CART process partitions the data and via equation 2 enables the each data pair membership grades to be calculated and fuzzified for application in the antecedent rules layer. In layer 2 the fuzzy rules are applied as shown in equation 3, determining the firing strengths of each antecedent rule.

$a_2 = \omega_i = \mu_{X1}(x_1) \cdot \mu_{X2}(x_2)$ for i = 1, …, m (3)

In the third layer the consequent rules are applied per Table 2, defuzzifying the output where a_i, b_i, c_i *for* $i = 1,\ldots, m$, are consequent coefficients ('c') and intercept of the linear function.

$a_3 = z_i = a_i + bx1_i + cx2_i$, *for* $i = 1, \ldots, m$ (4)

The results of the consequent functions are summed per equation 5 below, and provides the ANFIS predicted value.

$$a_4 = \sum_{j=1}^{n} a_2[j] \tag{5}$$

Table 2. Inference Rules to Consequent Functions Mapping.

		x_1	x_2	THEN				$x_1 \cdot$	$x_2 \cdot$
Rule #1	**IF**	**Cold**	**Dry**	**y**	**=**	a_{10}	**+**	b_{11}	c_{12}
Rule #2	**IF**	**Cold**	**Wet**	**y**	**=**	a_{20}	**+**	b_{21}	c_{22}
Rule #3	**IF**	**Hot**	**Dry**	**y**	**=**	a_{30}	**+**	b_{31}	c_{32}
Rule #4	**IF**	**Hot**	**Wet**	**y**	**=**	a_{40}	**+**	b_{41}	c_{42}

These predicted values are then compared to the actual dependent variable values to determine the overall error of the system. This study used the least squared error method and RMSE as the metric for prediction accuracy, see equation 1.

Another aspect of the CART ANFIS architecture is that the membership functions do not need to be refined as would be done in a general ANFIS model. This is not a relaxation of the structure, but taking advantage of the precision of the CART process on the front side in producing more accurate membership functions to start the processing and is implicit weight normalization.

The next step is to improve that result by tuning the 'c' values by the LSE and partial differential equations (PDE) in solving a system of multiple linear equations. This enables the 'c' values to learn and improve the values applied, therefore reducing the overall RMSE of the system.

The error function of the prediction output was defined where ε_i is the error of the i^{th} data point, shown in equation 6. Then the overall function (g) is compared with the actual dependent output (z), equation 7.

$$E = 1/2 \sum_{i=1}^{n} \varepsilon_i^2 \tag{6}$$

$$\varepsilon = g(x1_i, x2_i) - z_i = a + bx1_i + cx2_i - z_i \tag{7}$$

Combining equations 6 and 7, the PDEs are set up in order to determine the 'c' values. The PDEs needed are $\frac{\partial E}{\partial a}$, $\frac{\partial E}{\partial b}$, and $\frac{\partial E}{\partial c}$; setting each PDE to zero as a minimization problem.

These individual PDEs, equations 8–10, were the inputs to a multiple linear equation solving matrix as depicted in Table 3. When solved for each ‘c’ value, those were inserted back into Table 2 to update the ANFIS calculations as a one-step optimization of the ‘c’ values. Iterations of ‘c’ value optimization continue to reduce the RMSE, and therefore prediction accuracy.

$$\frac{\delta E}{\delta a} = 0 \rightarrow a\Sigma x1_i^2 + b\Sigma x1_i x2_i + c\Sigma x1_i - \Sigma x1_i z_i$$
$$\rightarrow a\Sigma x1_i^2 + b\Sigma x1_i x2_i + c\Sigma x1_i = \Sigma x1_i z_i \tag{8}$$

$$\frac{\delta E}{\delta b} = 0 \rightarrow a\Sigma x1_i x2_i + b\Sigma x2_i^2 + c\Sigma x2_i = \Sigma x2_i z_i \tag{9}$$

$$\frac{\delta E}{\delta c} = 0 \rightarrow a\Sigma x1_i + b\Sigma x2_i + nc = \Sigma z_i \tag{10}$$

Table 3. Multiple Linear Equations Solution Matrix Set-up.

		a_{10}	b_{11}	c_{12}	a_{20}	c_{21}	b_{22}	a_{30}	c_{31}	b_{32}	a_{40}	b_{41}	c_{42}	y
		1	2	3	4	5	6	7	8	9	10	11	12	13
a_{10}	1	Σd_1^2	$\Sigma d_1^2 x_1$	$\Sigma d_1^2 x_2$	$\Sigma d_1 d_2$	$\Sigma d_1 d_2 x_1$	$\Sigma d_1 d_2 x_2$	$\Sigma d_1 d_3$	$\Sigma d_1 d_3 x_1$	$\Sigma d_1 d_3 x_2$	$\Sigma d_1 d_4$	$\Sigma d_1 d_4 x_1$	$\Sigma d_1 d_4 x_2$	$\Sigma d_1 y$
b_{11}	2	$\Sigma d_1^2 x_1$	$\Sigma d_1^2 x_1^2$	$\Sigma d_1^2 x_1 x_2$	$\Sigma d_1 d_2 x_1$	$\Sigma d_1 d_2 x_1^2$	$\Sigma d_1 d_2 x_1 x_2$	$\Sigma d_1 d_3 x_1$	$\Sigma d_1 d_3 x_1^2$	$\Sigma d_1 d_3 x_1 x_2$	$\Sigma d_1 d_4 x_1$	$\Sigma d_1 d_4 x_1^2$	$\Sigma d_1 d_4 x_1 x_2$	$\Sigma d_1 x_1 y$
c_{12}	3	$\Sigma d_1^2 x_2$	$\Sigma d_1^2 x_1 x_2$	$\Sigma d_1^2 x_2^2$	$\Sigma d_1 d_2 x_2$	$\Sigma d_1 d_2 x_1 x_2$	$\Sigma d_1 d_2 x_2^2$	$\Sigma d_1 d_3 x_2$	$\Sigma d_1 d_3 x_1 x_2$	$\Sigma d_1 d_3 x_2^2$	$\Sigma d_1 d_4 x_2$	$\Sigma d_1 d_4 x_1 x_2$	$\Sigma d_1 d_4 x_2^2$	$\Sigma d_1 x_2 y$
a_{20}	4	$\Sigma d_1 d_2$	$\Sigma d_1 d_2 x_1$	$\Sigma d_1 d_2 x_2$	Σd_2^2	$\Sigma d_2^2 x_1$	$\Sigma d_2^2 x_2$	$\Sigma d_2 d_3$	$\Sigma d_2 d_3 x_1$	$\Sigma d_2 d_3 x_2$	$\Sigma d_2 d_4$	$\Sigma d_2 d_4 x_1$	$\Sigma d_2 d_4 x_2$	$\Sigma d_2 y$
b_{21}	5	$\Sigma d_1 d_2 x_1$	$\Sigma d_1 d_2 x_1^2$	$\Sigma d_1 d_2 x_1 x_2$	$\Sigma d_2^2 x_1$	$\Sigma d_2^2 x_1^2$	$\Sigma d_2^2 x_1 x_2$	$\Sigma d_2 d_3 x_1$	$\Sigma d_2 d_3 x_1^2$	$\Sigma d_2 d_3 x_1 x_2$	$\Sigma d_2 d_4 x_1$	$\Sigma d_2 d_4 x_1^2$	$\Sigma d_2 d_4 x_1 x_2$	$\Sigma d_2 x_1 y$
c_{22}	6	$\Sigma d_1 d_2 x_2$	$\Sigma d_1 d_2 x_1 x_2$	$\Sigma d_1 d_2 x_2^2$	$\Sigma d_2^2 x_2$	$\Sigma d_2^2 x_1 x_2$	$\Sigma d_2^2 x_2^2$	$\Sigma d_2 d_3 x_2$	$\Sigma d_2 d_3 x_1 x_2$	$\Sigma d_2 d_3 x_2^2$	$\Sigma d_2 d_4 x_2$	$\Sigma d_2 d_4 x_1 x_2$	$\Sigma d_2 d_4 x_2^2$	$\Sigma d_2 x_2 y$
a_{30}	7	$\Sigma d_1 d_3$	$\Sigma d_1 d_3 x_1$	$\Sigma d_1 d_3 x_2$	$\Sigma d_2 d_3$	$\Sigma d_2 d_3 x_1$	$\Sigma d_2 d_3 x_2$	Σd_3^2	$\Sigma d_3^2 x_1$	$\Sigma d_3^2 x_2$	$\Sigma d_3 d_4$	$\Sigma d_3 d_4 x_1$	$\Sigma d_3 d_4 x_2$	$\Sigma d_3 y$
b_{31}	8	$\Sigma d_1 d_3 x_1$	$\Sigma d_1 d_3 x_1^2$	$\Sigma d_1 d_3 x_1 x_2$	$\Sigma d_2 d_3 x_1$	$\Sigma d_2 d_3 x_1^2$	$\Sigma d_2 d_3 x_1 x_2$	$\Sigma d_3^2 x_1$	$\Sigma d_3^2 x_1^2$	$\Sigma d_3^2 x_1 x_2$	$\Sigma d_3 d_4 x_1$	$\Sigma d_3 d_4 x_1^2$	$\Sigma d_3 d_4 x_1 x_2$	$\Sigma d_3 x_1 y$
c_{32}	9	$\Sigma d_1 d_3 x_2$	$\Sigma d_1 d_3 x_1 x_2$	$\Sigma d_1 d_3 x_2^2$	$\Sigma d_2 d_3 x_2$	$\Sigma d_2 d_3 x_1 x_2$	$\Sigma d_2 d_3 x_2^2$	$\Sigma d_3^2 x_2$	$\Sigma d_3^2 x_1 x_2$	$\Sigma d_3^2 x_2^2$	$\Sigma d_3 d_4 x_2$	$\Sigma d_3 d_4 x_1 x_2$	$\Sigma d_3 d_4 x_2^2$	$\Sigma d_3 x_2 y$
a_{40}	10	$\Sigma d_1 d_4$	$\Sigma d_1 d_4 x_1$	$\Sigma d_1 d_4 x_2$	$\Sigma d_2 d_4$	$\Sigma d_2 d_4 x_1$	$\Sigma d_2 d_4 x_2$	$\Sigma d_3 d_4$	$\Sigma d_3 d_4 x_1$	$\Sigma d_3 d_4 x_2$	Σd_4^2	$\Sigma d_4^2 x_1$	$\Sigma d_4^2 x_2$	$\Sigma d_4 y$
b_{41}	11	$\Sigma d_1 d_4 x_1$	$\Sigma d_1 d_4 x_1^2$	$\Sigma d_1 d_4 x_1 x_2$	$\Sigma d_2 d_4 x_1$	$\Sigma d_2 d_4 x_1^2$	$\Sigma d_2 d_4 x_1 x_2$	$\Sigma d_3 d_4 x_1$	$\Sigma d_3 d_4 x_1^2$	$\Sigma d_3 d_4 x_1 x_2$	$\Sigma d_4^2 x_1$	$\Sigma d_4^2 x_1^2$	$\Sigma d_4^2 x_1 x_2$	$\Sigma d_4 x_1 y$
c_{42}	12	$\Sigma d_1 d_4 x_2$	$\Sigma d_1 d_4 x_1 x_2$	$\Sigma d_1 d_4 x_2^2$	$\Sigma d_2 d_4 x_2$	$\Sigma d_2 d_4 x_1 x_2$	$\Sigma d_2 d_4 x_2^2$	$\Sigma d_3 d_4 x_2$	$\Sigma d_3 d_4 x_1 x_2$	$\Sigma d_3 d_4 x_2^2$	$\Sigma d_4^2 x_2$	$\Sigma d_4^2 x_1 x_2$	$\Sigma d_4^2 x_2^2$	$\Sigma d_4 x_2 y$

IMPLEMENTATION – CART

The first computational step of the study was partitioning of the data to see which independent variable, temperature or rainfall was more influential on the results of the dependent variable, power. As shown in figures 1 and 2, temperature and rainfall vs. power charts, respectively, no clear split on either variable gives insight into what is the dominant one. Using CART, each cross-wise paring, temperature vs. power, and rainfall vs. power were analyzed to see where, and to what extent the error prediction of power was minimized. Tables 4 and 5, below, show the results leading to the choice of the first split, and therefore the root of the regression tree.

In Table 4, the calculations show where splits were made in calculating the minimal error for a partition with 40 degrees being the optimal split. The minimal error below 40 degrees with an error value of 126.6888 is the most accurate predictor of power use when dependent upon temperature. This is compared with the results of Table 5, showing the minimal error split at 2 inches of rainfall with an error value of 279.0056. Since 126.6888 is less than 279.0056, per the operations of CART, temperature is determined to be the root and first level of the regression tree as shown in Figure 1.

Table 4. Temperature vs. Power Error Partitioning

Temp	Count	$(\Sigma \text{ errors})^2$	Mean	Sqrt Mean Error
>=70	30	7411349.4593	247044.9820	497.0362
<70	60	3412671.6963	56877.8616	238.4908
>=60	42	8167452.8607	194463.1634	440.9798
<60	48	2656568.2948	55345.1728	235.2555
>=50	53	9016627.4751	170125.0467	412.4622
<50	37	1807393.6805	48848.4779	221.0169
>=40	78	10631420.6430	136300.2647	369.1887
<40	12	192600.5126	16050.0427	126.6888
>=30	87	10773124.2385	123829.0142	351.8935
<30	3	50896.9170	16965.6390	130.2522
			Min	126.6888

Table 5.Rainfall vs. Power Error Partitioning

Rainfall	Count	$(\Sigma \text{ errors})^2$	Mean	Sqrt Mean Error
>=6	12	3379318.646	281609.8872	530.6693
<6	78	7444702.51	95444.90397	308.9416
>=4.5	17	3935131.915	231478.3479	481.1220
<4.5	73	6888889.24	94368.34576	307.1943
>=3	47	6301360.063	134071.4907	366.1577
<3	43	4522661.092	105178.1649	324.3118
>=2	73	9500670.796	130146.1753	360.7578
<2	17	1323350.36	77844.13879	279.0056
>=1	84	10276672.88	122341.3438	349.7733
<1	6	547348.2785	91224.71309	302.0343
			Min	279.0056

Setting temperature as the root independent variable, each of the subsequent branches on level 2 were analyzed in the same manner for rainfall vs. power to determine best split respective for both less than, and greater than / equal to 40 degrees. The left branch was investigated first towards refining the overall accuracy. Here an absolute reduction of error was found, minimizing error to 0 when the rainfall was greater than 6 inches, as shown in Table 6 as Min. 1. When cross referenced with the data set, no instances of rainfall greater than 6 inches occurred when the temperature was below 40 degrees. With this null set identified, the next larger error value was for a split at 3 inches of rainfall, as determined with an error value of 106.4875, Min.2 in Table 6. These splits are depicted as levels 3, 6 inches, and 4, 3 inches, on Figure 1. No further splits were done on the left side of the regression tree.

The analysis continued on the right side of the tree to determine the split for level 3. The focus in determining the minimum error value for rainfall for temperatures greater than / equal to 40 degrees. The best split was determined in last calculation with 1 inch of rain being the discriminating value. The best error was 278.2475, as shown in Table 7. Therefore the final partition of the overall data set completing the study's regression tree in Figure 1. Another representation of how the data set was partitioned out is shown in Figure 3, color coded to represent the area with the null, white area of influence (1) that had no data points, the best, dark green (3) with the most reduced estimation error, next best, light green (2) to poor, yellow (5), and worst, red, estimation (4).

Table 6. Rainfall Regression Partitioning (Temperature <40°).

For Temp <40 Rainfall	Count	(Σ errors)^2	Mean	Sqrt Mean Error
>=6	0	0.0000	0.0000	0.0000
<6	12	192600.5126	16050.0427	126.6888
>=4.5	1	15920.8316	15920.8316	126.1778
<4.5	11	176679.6810	16061.7892	126.7351
>=3	6	68037.5007	11339.5835	106.4875
<3	6	124563.0119	20760.5020	144.0851
>=2	9	156920.1067	17435.5674	132.0438
<2	3	35680.4059	11893.4686	109.0572
>=1	11	162609.9699	14782.7245	121.5842
<1	1	29990.5427	29990.5427	173.1778
			Min 1	0.0000
			Min 2	106.4875

Table 7. Rainfall Regression Partitioning (Temperature >=40°).

Temp >=40 Rainfall	Count	(Σ errors)^2	Mean	Sqrt Mean Error
>=6	12	3379318.6459	281609.8872	530.6693
<6	66	7252101.9970	109880.3333	331.4820
>=4.5	16	3919211.0835	244950.6927	494.9249
<4.5	62	6712209.5595	108261.4445	329.0311
>=3	41	6233322.5625	152032.2576	389.9131
<3	37	4398098.0805	118867.5157	344.7717
>=2	64	9343750.6894	145996.1045	382.0944
<2	14	1287669.9536	91976.4253	303.2762
>=1	73	10114062.9072	138548.8069	372.2214
<1	5	387108.3393	77421.6679	278.2475
			Min	278.2475

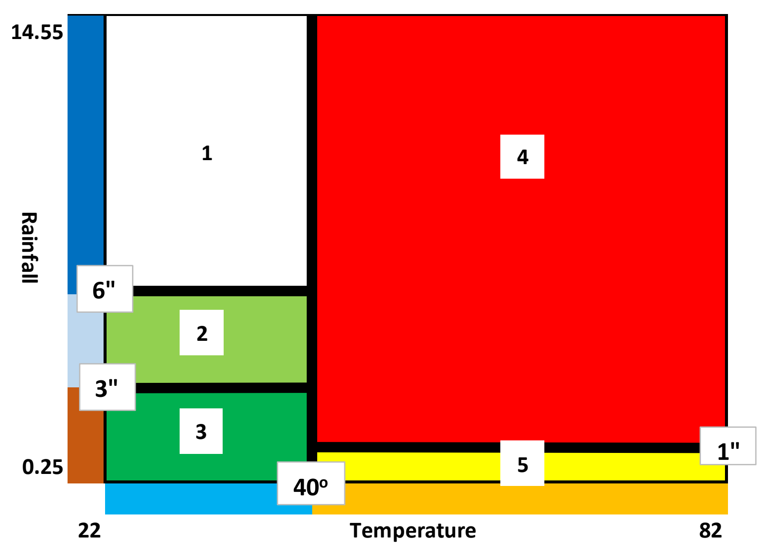


Figure 3. Data Set Partitioning to Reduce Estimation Error.

IMPLEMENTATION – ANFIS

The results of the CART regression analysis prepared the data to be used for processing through the ANFIS algorithm. Setting up the CART results as membership functions was the next step taken. This entailed identifying appropriate initial values for α, β, and γ for use in equation 2 in plotting out the membership functions as shown in Figures 4-6. The γ parameters were the first to be assigned at the endpoint values across all data set categories: 22 and 82 for temperature, 0.25 and 6 for rainfall vs. cold, 0.25 and 14.55 for rainfall vs. hot. The next parameters identified were the α values. This was accomplished by adding/subtracting the CART split point from both the high and low γ values. This point was also the cross-over point within the membership function, producing the following α values: 18 for temperature: cold; 42 for temperature:hot; 3.25 and 3.00 for cold:dry and cold:wet respectively; and finally 1.25 and 13.55 for hot:dry and hot:wet respectively. These parameters are shown in Table 8 below. The final step of preparing the membership functions was determining appropriate β values. The goal in establishing appropriate β values was ensuring that the values of the membership grades for each data point summed to one. Therefore, the β was key in determining this. The LSE method was applied here, focused on setting the β values with minimal cumulative deviance from the sum to 1 target. The cumulative deviation values were reduced to 0.0151 for temperature, and 0.1162 and 0.0395 for rainfall below and above the 40 degree temperature split, respectively. These values produced the shape for each membership function. Additionally, developing the membership functions from CART and by reducing the sum to deviations within each membership function, further tuning of them across the ANFIS algorithm was unnecessary; and, the removal of the standard ANFIS normalization layer(s) was acceptable.

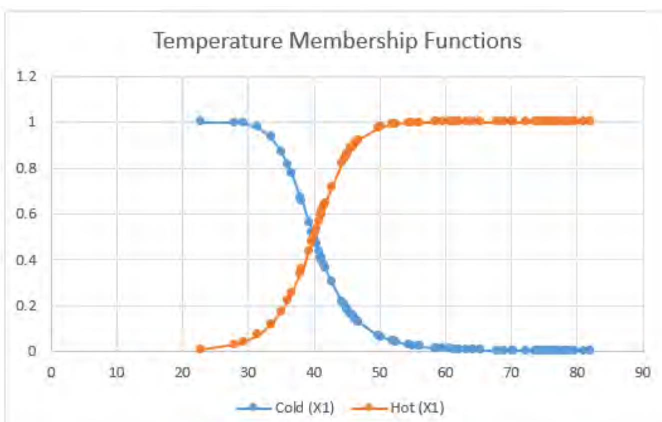


Figure 4. Temperature Membership Functions (Cold vs. Hot).

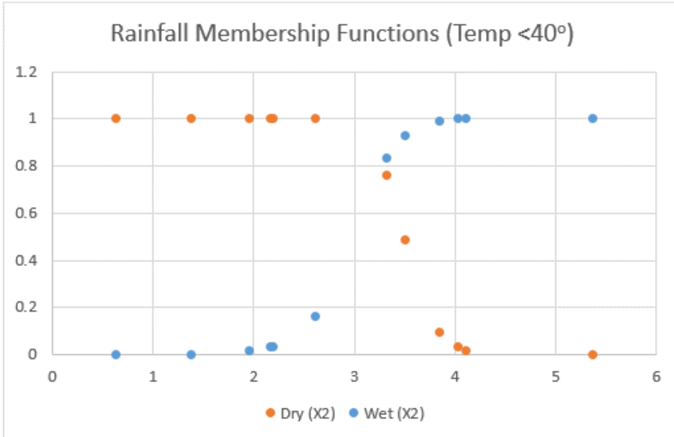


Figure 5. Rainfall Membership Functions (Dry vs. Wet; Temperature <40°).

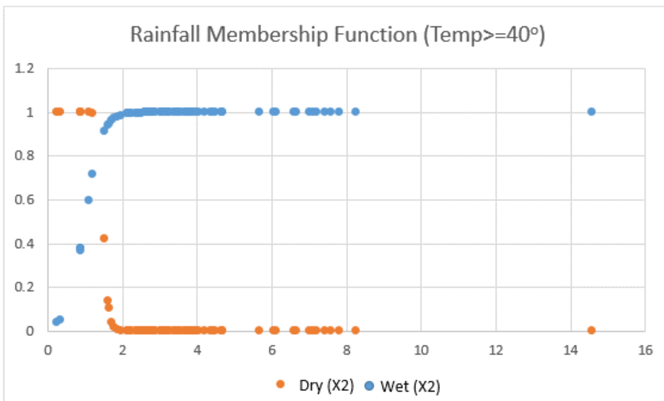


Figure 6. Rainfall Membership Functions (Dry vs. Wet; Temperature >=40°).

Table 8. Membership Function Parameters.

		α	β	γ	
Rainfall (X_1)	Dry (<40°)	3.2500	11.0000	0.2500	(<3")
	Wet (<40°)	3.0000	7.0000	6.0000	(>=3")
	Dry (>=40°)	1.2500	10.0000	0.2500	(<1")
	Wet (>=40°)	13.5500	30.0000	14.5500	(>=1")
Temp. (X_2)	Cold (<40°)	18.0000	3.0000	22.0000	
	Hot (>=40°)	42.0000	7.0000	82.0000	

The next step in building the ANFIS algorithm was calculating the membership grades in layer 1 of Fig. 2, applying the equation 2 to all the data points. This is where the next ANFIS alteration was implemented due to using CART in order to set up the membership functions. As shown in Fig. 2, in layer 1, there is an inverse node downstream of the membership function node ($1-\mu_{x1<40\ deg.}$). This inverse node permits a consolidated number of antecedent rules. In a standard fuzzy inference system the number of rules is relational total across the membership functions. In the present case that would be 2 x 2 = 4 for the temperature applied to the <40 rainfall data set and a 2 x 2 = 4 rule set for the temperature >= 40 rainfall data set. Because the <40 and >=40 data sets are mutually exclusive, the inverse values are applied to the opposite calculations. With this consolidation, the rule set as shown in Table 9, 4 rules, were used in the study. Each membership grade is used to determine the weights, ‘c’ parameters, for that rule in order to improve accuracy implementing the four equations associated with the rules in Table 1.

The rest of the ANFIS algorithm was executed as described above, determining RMSE from the comparison of the calculated output estimates against the actual dependent data points. The ‘c’ parameters were then tuned by LSE using the vectors for each of the independent variables, dependent variable, and the membership grades of the independent variables. The PDE matrix was established for each of the ‘c’ parameters, 12 total, a,b,c for each of the 4 rules, in refining the values and lowering RMSE.

Table 9. Final Trained Rule Consequent Parameters, RMSE = 278.

Rule / 'c'	a	b	c
1	435.2596	10.7509	40.0182
2	-129.7446	12.5157	6.9074
3	705.1781	4.7385	-219.7777
4	468.0829	6.0760	17.8385

DESCRIPTION OF RESULTS

Being able to join the computational capabilities for both CART and ANFIS into one structure, getting acceptable predictions was further enhanced with the transparency of the processes. Reducing the overall power RMSE to 278 from the initial settings output of 958, was a 70% reduction in error. An intermediate, manual tuning was able to reduce the RMSE to 341, and solving the multiple linear equations matrix proved to be effective in further improving the predictive accuracy of the hybrid algorithm. An effective mapping from the chosen independent variables to an accurate dependent variable prediction was shown in the results.

The process by which the CART regression broke out the most influential sections of the data made intuitive sense, added clarity to the establishment of the linguistic variable interpretation, and made the development of the membership functions straightforward. The combination of transparency and accuracy was useful in understanding the mechanisms used

CONCLUSIONS AND DISCUSSION

This study was successful in combining the CART method with ANFIS, producing consistent results that were tunable. The output of the CART processing aided in both setting up the membership functions effectively and increasing transparency in creating the linguistic variables, bridging the understanding gap between the user and the mathematical algorithms. Additionally, using CART decreased the complexity of the ANFIS algorithm three observed ways: compactness (removal of a layer of computation, implicit normalization), increasing simplicity (allowed for four rules to be set up and used with the partitions presented via the inverse function for membership grades), and further simplification from the standard ANFIS in that the membership functions need not be further tuned due to the pre-loading work of CART methodology as the input to ANFIS processing.

The presented results offer an introduction to a powerful methodology towards more in-depth use and applications of CART and ANFIS capabilities. Through implementing the historical

CART as the front-end data structuring process with the capabilities of ANFIS as a hybrid computing method offers another option in forecasting across multiple domains. CART refines estimating the number of membership functions and the number of IF-THEN reasoning rules needed for ANFIS to function effectively and more efficiently than only relying on gradient descent and least squares methods. Using CART also relieves the need to have a normalization layer as part of the ANFIS architecture. (Jang, 1994) The transparency offered by the hybrid implementation of enables the practitioner and the decision-maker the opportunity to understand the processes in following the logic. The applicability of fuzzy logic to human inexact reasoning can add to developing social and organizational models from which policy makers and gain insight into human reasoning. (Kar, et.al, 2014) With this deeper understanding, both professionals can explore possible futures and test hypothesis more readily and greater clarity between options, or across a continuous spectrum, due to the fuzzy nature of the calculus fashioned towards imitating human inexact reasoning.

Additional areas for extending this investigation include revisiting all the data categories collected to determine which has the most influence on power use; further partitioning of the data-space via CART regression to better refine the predictive capabilities; exploring changes needed to make the model applicable to other neighborhoods and demographics – and simplifying the tailoring to help other consumers use the same structure, and accurate to specific households; and further analysis on the cost of the power used to see the effects of inflation on the predictive power of the CART ANFIS model.

ACKNOWLEDGEMENTS

The author wishes to express sincere gratitude to the anonymous reviewers for their constructive comments and helpful suggestions which lead to substantial improvements.

References

Breiman, L. (2017). Classification and regression trees. Routledge.

Fragiadakis, N. G., Tsoukalas, V. D., & Papazoglou, V. J. (2014). An adaptive neuro-fuzzy inference system (anfis) model for assessing occupational risk in the shipbuilding industry. Safety Science, 63, 226-235.

Jang, J. S. R. (1991, July). Fuzzy modeling using generalized neural networks and Kalman filter algorithm. In *AAAI* (Vol. 91, pp. 762-767).

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE transactions on systems, man, and cybernetics, 23(3), 665-685.

Jang, J. S. (1994, June). Structure determination in fuzzy modeling: a fuzzy CART approach. In *Proceedings of 1994 IEEE 3rd International Fuzzy Systems Conference* (pp. 480-485). IEEE.

Jang, J. S. (1996, September). Input selection for ANFIS learning. In *Proceedings of IEEE 5th international fuzzy systems* (Vol. 2, pp. 1493-1499). IEEE.

Jang, J.R., C-T. Sun, and E. Mizutani (1997). Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. Upper Saddle River, NJ: Prentice Hall.

Kar, S., Das, S., & Ghosh, P. K. (2014). Applications of neuro fuzzy systems: A brief review and future outline. Applied Soft Computing, 15, 243-259.

National Oceanagraphic and Atmospheric Administration. web page (accessed 30 September, 2020):https://www.ncdc.noaa.gov/IPS/cd/cd.html;jsessionid=7F5D8CA98EFF7E639E3E32486C2F0BD8?_page=0&jsessionid=7F5D8CA98EFF7E639E3E32486C2F0BD8&state=MD&_target1=Next+%3E

Olaru, C. and Wehenkel, L. (2003). A Complete Fuzzy Decision Tree Technique. Fuzzy Sets and Systems, Elsevier.

Pinem, A. F. A., & Setiawan, E. B. (2015, May). Implementation of classification and regression tree (CART) and fuzzy logic algorithm for intrusion detection system. In 2015 3rd International Conference on Information and Communication Technology (ICoICT) (pp. 266-271). IEEE.

Poor Arab Moghadam, M., Pahlavani, P., & Naseralavi, S. (2016). Prediction of car following behavior based on the instantaneous reaction time using an ANFIS-CART based model. International Journal of Transportation Engineering, 4(2), 109-126.

Raja, P., & Pahat, B. (2016). A review of training methods of ANFIS for applications in business and economics. International Journal of u-and e-Service, Science and Technology, 9(7), 165-172.

Southern Maryland Electric Cooperative website, accessed 30 March 2021: https://smeco.coop/

Weather Underground, Patuxent River, MD Weather History website (accessed 30 March, 2021):https://www.wunderground.com/history/monthly/us/md/patuxent-river/KNHK/date/2011-8

Yang, B. S., Oh, M. S., & Tan, A. C. C. (2009). Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(2), 1840-1849.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. IEEE Transactions on systems, Man, and Cybernetics, (1), 28-44.

Zadeh, L. A. (1994). Soft computing and fuzzy logic. IEEE software, 11(6), 48-56.

DECISION SCIENCES INSTITUTE

Meta-learning of Algorithm Selection for Visual Inspection through Case-based Reasoning

Shu Wang
Georgia Institute of Technology
Email: swang737@gatech.edu

Mulang Song
Georgia Institute of Technology
Email: msong84@gatech.edu

Yiyun (Cindy) Fei
Georgia Institute of Technology
Email: yfei38@gatech.edu

Jianxin (Roger) Jiao
Georgia Institute of Technology
Email: rjiao@gatech.edu

ABSTRACT

Visual inspection is an important task in manufacturing automation to improve product quality and reduce cost of poor quality. Because it happens in different procedures and products, defects can have unique features, and solutions vary a lot. This makes inspection algorithm selection a challenging question, and the reasoning requires much engineering experience. To assist the decision-making about which algorithm is likely to perform well for an inspection task, this paper proposes a meta-learning approach for learning-to-learn from historical projects. An improved case-based reasoning approach with knowledge graph embedding is proposed for meta-learning. An example case is demonstrated to illustrate it.

KEYWORDS: Meta-learning, Knowledge Graph, Graph Embedding, Case-based Reasoning, Visual Inspection

INTRODUCTION

For manufacturing automation, visual inspection is to use machine vision and machine learning algorithms to conduct defect inspection for different products in multiple stages, which helps manufacturers reduce cost of labor and increase yield and quality. Because inspection happens on a case-by-case basis and there is no one algorithm that can solve all problems, the vision solutions shall be proposed to address the unique features for each case. For instance, cosmetic defect detection systems can use algorithms like blob detection and YOLO neural networks, while measurement inspection systems need line detection or circle detection algorithms. Due to the uniqueness of each inspection task, algorithm selection, which is to find which algorithm is likely to perform well for this task (Smith-Miles, K. A., 2009), requires much engineering experience and is a highly manual process that involves lots of trials and errors.

To solve the algorithm selection problem for visual inspection, this paper proposes a meta-learning approach to enable learning-to-learn of inspection solution development as a decision support agent, which allows the system to recommend algorithms that potentially solve a new inspection task. By learning experience from past inspection projects, the meta-learning system can identify one promising feature-classifier combination that fits a new inspection task with high performance, thus facilitating the solution development and making the process less rely on the engineer's experience.

There are several challenges for implementing meta-learning of algorithm selection. (i) Inspection case representation. Meta-learning aims to learn how to effectively solve a new inspection task from past visual inspection cases. How to effectively represent a case using available features is an important question for follow-up activities. (ii) Recommendation mechanism. Algorithm selection is to predict which algorithm suits a new inspection task, and how to find the algorithm that has good performance for the new task needs to be answered. (iii) Historical case maintenance. Maintenance for the past cases is a challenging question for meta-learning. As cases accumulate, newer inspection tasks may involve features that old cases do not consider. Therefore, maintenance shall consider the scalability of the case base.

To address the above challenges, this paper proposes a knowledge graph embedding-based case-based reasoning approach for meta-learning of algorithm selection. By using knowledge graphs, cases in the form of tables can be embedded into a low-dimensional space through embedding operations. In this way, algorithm recommendation can be conducted by finding a similar case in the embedding space and then adapting the old solution. Furthermore, maintenance of inspection cases becomes easier, and adding a new case feature can be completed by adding new types of relations and entities.

The rest of this paper is organized as follows. Literature review introduces the previous work in meta-learning, symbolic reasoning, and knowledge graph. Then the formulation of case-based reasoning for meta-learning is presented. After that, the proposed knowledge graph embedding approach for case representation is introduced. An example case is presented to validate the feasibility of the proposed meta-learning approach of algorithm selection. Finally, the managerial implications and concluding remarks are made.

LITERATURE REVIEW

Meta-learning

Deploying different machine learning techniques in real-world projects usually rely on hand-crafting solutions by experienced practitioners, and meta-learning (or learning-to-learn), tries to replace human insights by systematically learning how to learn more efficiently based on prior experience with other tasks (Lansdell & Kording, 2019). There are two streams of research about meta-learning, one is to study how to determine efficient optimization procedures, and the other is to make an existing agent quickly generalized to new tasks. The objective of the former stream is to characterize tasks with meta-data and to build meta-models that learn the mapping between task characteristics and learning performance, while the latter is to study how to transfer learned models to new tasks that are inherently similar (Vanschoren, 2018). For learning to optimize tasks, there are studies in meta-feature selection (Rivolli et al., 2018) and meta-model development, including model building for algorithm selection (Bischl et al., 2016) and configuration or hyperparameter recommendation (Ali & Smith-Miles, 2006). For structure learning, the main study topics include transfer learning (Weiss et al., 2016) and few-shot learning, where each classification task has very few images but there are enough datasets from different tasks (Lake et al., 2017).

Symbolic Reasoning

Looking for a proper algorithm for a new visual inspection task is a decision-making process that is based on prior experience and knowledge. Symbolic reasoning mimics the human decision-making process to help inference automation. Prevailing symbolic reasoning approaches include fuzzy logic, rough set theory, case-based reasoning, and so on. Fuzzy set theory is a theoretical approach to ambiguity (Suraj, 2004), and rough set theory is also a mathematical approach to reasoning with ambiguity and imperfect knowledge (Pawlak, 1998). It is helpful when there are nominal factors in extracted features. Case-based reasoning does not require modeling of the domain knowledge, but it solves new problems by adapting previous successful solutions from historical similar cases (Watson, 1994).

Knowledge Graph

Knowledge graphs are used to construct and describe large entity-relation networks and can represent knowledge bases in the form of graph structure (Chen et al., 2020), where entity types are represented with nodes and relations with edges. Common reasoning tasks for knowledge graphs include node classification, link prediction, graph classification, clustering, and predictive queries. Node classification and link prediction focus on a single entity and relationship in an incomplete knowledge graph, and graph classification is at the graph level.

For the defect detection domain, potential applications include clustering and predictive query tasks. Clustering is to determine if nodes form a community, which can be used to cluster the subjectively defined defect types based on feature nodes. This can provide insights in the defect feature level to study these defects. For query tasks, the objective is to search entities based on query conditions regarding entity and relation types. Usually, the knowledge graph is not directly used for reasoning, but the graph needs to be embedded into a lower-dimensional space, where both nodes and edges have their embedding functions. In this way, queries are turned into calculatable operations. Prevailing embedding methods include TransE (Bordes et al., 2013), TransR (Lin et al., 2015), DistMult (Yang et al., 2014), and ComplEx (Trouillon et al., 2016).

CASE-BASED REASONING FOR META-LEARNING

Case-based reasoning is a tool to reference historical examples for decision-making and can be used for infrequent reasoning activities that rely on human experience. The advantage over conventional rule-based reasoning is that it does not need expert knowledge to develop a knowledge base. Instead, it relies on the accumulation of solved cases and the ability to identify similar historical cases. Then the solution to the current case is the reuse or modification of the historical solution. This enables the system to evolve as new cases continuously accumulate.

Generally, the implementation of case-based reasoning involves the following steps: (i) case representation and indexing, (ii) new case creation, (iii) case retrieval, (iv) case reuse or case adaptation, and (iv) case retention. Case representation and indexing is to organize the cases in a systematic format for follow-up processing. Since case-based reasoning relies on knowledge from past cases, a case library is developed to keep the solved cases. The scalability and generalization of case representation and the ease of case search are two targets during case library development. After the development, a new case can be initialized and indexed using the same format as that in the case library, and case retrieval is to conduct similarity search for the most similar historical case with the new one. The similarity measurement approach is usually predefined, which uses a similarity measure function that is designed subjectively. One common approach is to use the weighted sum of different case attributes. After finding the most similar historical case, case reuse or adaptation is conducted to decide the new solution based on the historical solution. Production rules can be used here to indicate how the case should be adapted based on the difference between the two cases. When the solution is deployed, feedback is collected for the new case on how the adapted solution works, and the new case will be stored in the library for future reference.

Case-based reasoning can be applied and beneficial to manufacturing activities. In manufacturing industries, successful inspection automation requires large efforts of domain experts and engineers. Each inspection project can have unique defect features and require different machine vision techniques, therefore deploying a machine vision solution from scratch takes much time. Since there is not such an algorithm that can fit all scenarios with good performance, automating the process of algorithm selection is beneficial to practitioners. The objective of meta-learning is to conduct automated machine learning algorithm selection through meta-features of past learning cases (Garouani et al., 2022). For an automation department in a manufacturing company, visual inspection tasks are widely conducted in different product lines. If these tasks can be accumulated and learned to provide guidance for new inspection tasks through case-based reasoning, practitioners can depend less on expert experience regarding algorithm selection and configuration recommendation.

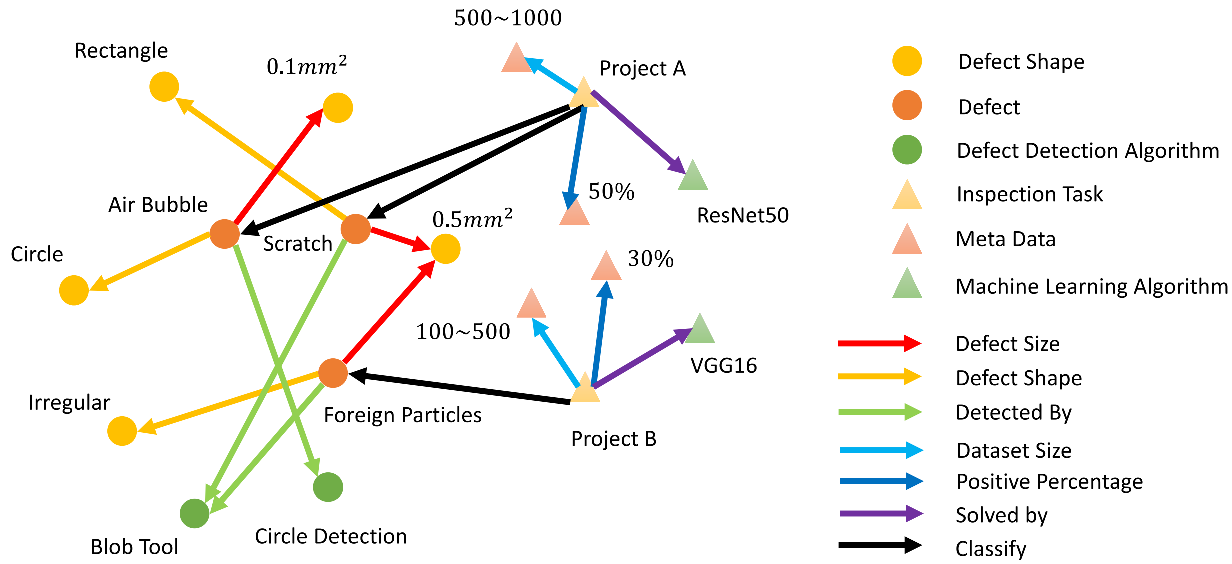
To implement meta-learning of algorithm selection, case-based reasoning provides a new perspective. Each case can be modeled as an inspection project, and meta-features and target defects from each project can be organized as the case features. After the attributes for case representation are determined, a case library can be developed. It stores historical projects with their meta-features and algorithm performance. When a new inspection project begins, a new case can be created based on the available meta-features. Case retrieval is then conducted to identify the most similar case in the library, and the algorithm selected from that case can be reused or modified based on the difference of case features.

KNOWLEDGE GRAPH EMBEDDING FOR CASE REPRESENTATION AND RETRIEVAL

To enable case-based reasoning for meta-learning, this paper proposes a knowledge graph embedding approach for case representation and retrieval. Compared to the conventional symbolic case representation (Ali et al., 2018), the proposed approach has several advantages. (i) It adds defect descriptions as a new dimension other than meta-features. Meta-learning aims to learn the relation mapping between meta-features and algorithm performance. However, it lacks descriptions of the problem. In the visual inspection domain, defect features are important factors for algorithm selection. (ii) It has better scalability. Original case representation requires an additional attribute for a new case feature. In the knowledge graph, this can be done by adding a new type of relations, which makes it easier to manage the case library. (iii) It naturally allows semantic feature descriptions and calculations, making it easier for case retrieval. Semantic descriptions can be initialized as entity nodes and get embedded in the embedding space for difference calculation. (iv) Knowledge graphs can express relations between case features, thus visualizing case similarities. Plus, the similarity is not manually designed like traditional case-based reasoning but through minimizing the loss function during triple training.

For visual inspection tasks, meta-features and defect features are both important for algorithm selection. Meta-features include the number of data, the number of classes, the number of positive samples, and some other attributes that describe the dataset. To represent a visual inspection case in a knowledge graph, several types of entity nodes can be used. (i) Defect features. Defect features can use both geometry and color information. For example, (*blue, background color, defect1*) can be used to represent defect background color, and (*circle, shape, defect1*) can represent the defect shape. (ii) Defect instances. This entity type represents a defect type. (iii) Defect detection algorithms. This refers to image processing or machine vision algorithms for defect detection. For instance, (*blob tool, detect, defect1*) represents the defect1 can be detected by the blob tool. (iv) Inspection tasks. An inspection task can contain one or multiple defect types, while the task itself can be classification or object detection. For example, two triples, (*task1, classify, defect1*) and (*task1, classify, defect2*), represent the task1 is to classify the defect1 and the defect2. (v) Meta-features of inspection tasks. Meta-features refer to the attributes of the dataset. In the proposed approach, meta-features are nominal instead of numeric, where the discretization is based on expert knowledge. For example, ($100{\sim}500$, *dataset size, task1*) can be used to represent the number of data points for the task1. (vi) Classification algorithms. This refers to the machine learning algorithms used for a classification task. For instance, (*task1, solved by, ResNet*) means the task1 is addressed by the ResNet algorithm. Figure 1 is an example of a knowledge graph for inspection case representation.

Figure 1. An Illustrative Example of Visual Inspection Knowledge Graph



Using knowledge graphs for case representation require different methods for implementing case representation and retrieval compared to conventional approaches. Different from using the weighted sum of various attribute differences, case similarity in knowledge graphs can be estimated through the distance in the embedding space, and it is learned through the loss function with head-relation-tail structures during embedding model training. Therefore, when a new inspection project is initiated, a new case can be created and embedded into the entity space, and the closest case entity can be found. Case adaptation or reuse can then be conducted. After the solution is deployed, the original missing relation with the inspection algorithm can be completed, and the new knowledge graph embeddings can be updated for retention of this new case.

As previously discussed, knowledge graph embedding is used for projecting case entities into the embedding space, so that case retrieval can be conducted by initiating a new case and finding the closest case. In this proposed framework, the TransR embedding approach is selected. There are several considerations. (i) The visual inspection case knowledge graph shall model antisymmetric relations, inverse relations, and one-to-many relations. Antisymmetric relation means that the relation is not bidirectional, like one case is solved by one algorithm. Inverse relation refers to a pair of relations between heads and tails, like “is solved by” and “solves” between a case and an algorithm. One-to-many relations mean that there can be multiple tails going from one head through one relation. For instance, a case can classify multiple types of defects. Table 1 lists the expressiveness of prevailing embedding approaches. (ii) The relations in the visual inspection case knowledge graph are simple and translation-based, and TransR is sufficient to support such relation embeddings. (iii) Because TransR embeddings are real values, its embeddings are more straightforward and interpretable compared to ComplEx. (iv) Both TransR and ComplEx requires additional techniques for two-dimensional visualization. TransR embeddings can be transformed from high-dimensional space to two-dimensional space using principal component analysis (PCA) or t-SNE (Van der Maaten & Hinton, 2008), while ComplEx embeddings should be converted from a complex vector space to a real vector space.

Table 1. The expressiveness of Embedding Models

Model	Embedding Space	Symmetric Relation	Antisymmetric Relation	Inverse Relation	Composite Relation	One-to-Many
TransE	$\mathbb{R}$		×	×	×	
TransR	$\mathbb{R}$	×	×	×		×
DistMult	$\mathbb{R}$	×				×
ComplEx	$\mathbb{C}$	×	×	×		×

For TransR embeddings, entities and relations are modeled in two distinct spaces, i.e., entity space and relation space, and the reasoning of the tail entity is through the translation of the head entity in the relation space, or vice versa (Lin et al., 2015). Each relation is modeled to have a distinct interaction pattern between entities.

For each triple (h, r, t), where h and t are head and tail entities, and r is the relation, entity embeddings are set as $\mathrm{h}, \mathrm{t} \in \mathbb{R}^k$ and relation embeddings as $\mathrm{r} \in \mathbb{R}^d$. A project matrix $M_r \in \mathbb{R}^{k \times d}$ is set to project entities from entity space to relation space:

$$\mathrm{h}_r = \mathrm{hM_r}, \mathrm{t}_r = \mathrm{tM_r} \quad (1)$$

The score function for each triple is defined accordingly:

$$f_r(h,t) = \|\mathrm{h}_r + r - t_r\|_2^2 \quad (2)$$

During training, the model uses a margin-based ranking loss and aims to minimize the score function for true triples while maximizing it for incorrect triples:

$$L = \sum_{(h,r,t)\in S} \sum_{(h',r',s')\in S'} \max\big(0, f_r(h,t) + \gamma - f_r(h',t')\big) \quad (3)$$

where γ is the margin, and S and S' represent the set of correct triples and incorrect triples.

EXAMPLE CASE

In this paper, a case library with 51 visual inspection cases is developed from real-world visual inspection projects. A new project of catheter visual inspection in an assembly line is taken as an example to demonstrate knowledge graph embedding-based case-based reasoning for meta learning. Catheters are one common product in medical industries. Because medical products require high quality, inspection needs to identify tiny defects. The ambiguous geometric features, variety in defect types, and the small size of defects make the inspection a very manual process and difficult to automate. During the solution development, knowledge graph embedding-based case-based reasoning is utilized.

There are two types of defects for classification in this case. The camera captures monochrome images for both types of defect samples. One type of defect has a circle shape, and the color is gray. The size is tiny. The other type of defect has a strip feature, and its size is small. Its aspect ratio is usually larger than 10. There are also some positive samples that have slightly similar features with the two types of defects. This dataset contains 4,000 images, and the positive samples occupy about 50% of the dataset. Using the above information, a new case can be initiated in the knowledge graph by inserting triples displayed in Table 2. Then a new model is trained to learn the embeddings of the updated graph, and the training curve is displayed in Figure 2. After updating the embeddings for this new case, the nodes are visualized in two-dimensional space using the PCA. The entity plot and relation plot are shown in Figure 3 and Figure 4. The closest case to the newly created case is found, and its solution is retrieved, suggesting the classification could be conducted using the VGG16 algorithm. After the model training using VGG16, the testing dataset achieves an accuracy that is above 98%, which validates the effectiveness of the proposed meta-learning approach.

Table 2. Triples related with the new inspection case

Case52	hasColor	Monochrome
Case52	task	Multiclassification
Case52	numberOfImagePerClass	Large
Case52	positivePercentage	Even
Case52	classify	Defect347
Case52	classify	Defect348
Defect347	hasShape	Circle
Defect347	defectSize	Tiny
Defect348	hasShape	Rectangle
Defect348	defectSize	small

Figure 2. Loss Curve of the Embedding Model Training

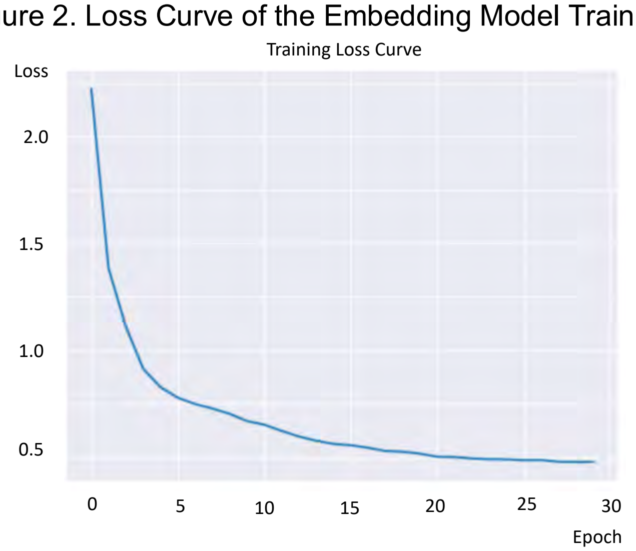


Figure 3. Entity Plot

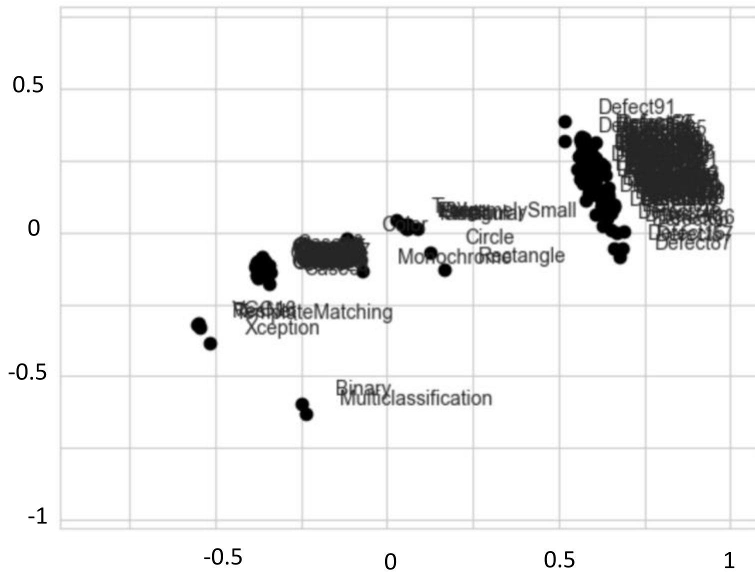
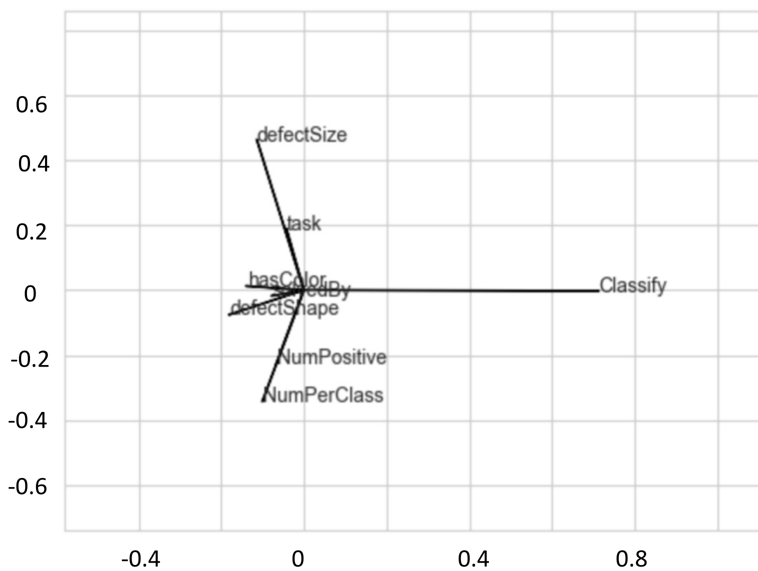


Figure 4. Relation Plot



MANAGERIAL IMPLICATIONS

In this paper, a knowledge graph embedding-based case-based reasoning is proposed for meta-learning of visual inspection algorithm selection. It has significant practical benefits. Compared to the prevailing machine vision system development process, the proposed case-based reasoning approach for meta-learning can be very helpful for companies to accumulate inspection cases. Since companies can learn from their own projects, the case base is more specialized towards the company products. This can save many engineering resources and reduce the solution development time. Therefore, the proposed meta-learning approach should help companies to win more customers by providing high-performance solutions with a shorter lead time.

CONCLUDING REMARKS

This paper proposes a case-based reasoning approach with knowledge graph embedding for meta-learning of visual inspection algorithm selection. The contributions of this paper contain several points. Firstly, it proposes a case-based reasoning approach for meta-learning of visual inspection algorithm selection. It addresses the lengthy algorithm selection process that requires large efforts and experience, thus shortening the development duration with an efficient knowledge management and inference approach. Secondly, it proposes an improved case-based reasoning approach by using knowledge graph embedding for case representation and retrieval. This increases the usability and inference capability of case-based reasoning.

There are several limitations for the existing work. Firstly, as a symbolic reasoning approach, case-based reasoning has limitations in capturing the underlying numerical relationships between meta-features and algorithm performance. Secondly, meta-features are embedded as nominal factors in the knowledge graph. This makes it better at handling semantic processing tasks, but it can have difficulties reflecting numeric values of the non-nominal features. Thirdly, the proposed case similarity measurement approach through knowledge graph embedding is influenced by the embedding quality. Therefore, to support the usage scenarios where the number of available data is small, how to train an embedding model with limited number of training data should be studied.

REFERENCE

Ali, R., Khatak, A. M., Chow, F., & Lee, S. (2018). A case-based meta-learning and reasoning framework for classifiers selection. In Proceedings of the 12th international conference on ubiquitous information management and communication, January, 1-6.

Ali, S., & Smith-Miles, K. A. (2006). A meta-learning approach to automatic kernel selection for support vector machines. Neurocomputing, 70(1-3), December, 173-186.

Bischl, B., Kerschke, P., Kotthoff, L., Lindauer, M., Malitsky, Y., Fréchette, A., Hoos, H., Hutter, F., Leyton-Brown, K., Tierney, K. and Vanschoren, J. (2016). Aslib: A benchmark library for algorithm selection. Artificial Intelligence, 237, August, 41-58.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. Advances in neural information processing systems, December, 26.

Chen, X., Jia, S., & Xiang, Y. (2020). A review: Knowledge reasoning over knowledge graph. Expert Systems with Applications, 141, March, 112948.

Garouani, M., Ahmad, A., Bouneffa, M., Hamlich, M., Bourguin, G., & Lewandowski, A. (2022). Using meta-learning for automated algorithms selection and configuration: an experimental framework for industrial big data. Journal of Big Data, 9(1), April, 57.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and brain sciences, 40.

Lansdell, B. J., & Kording, K. P. (2019). Towards learning-to-learn. Current Opinion in Behavioral Sciences, 29, October, 45-50.

Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In Twenty-ninth AAAI conference on artificial intelligence, 29(1), February.

Pawlak, Z. (1998). Rough set theory and its applications to data analysis. *Cybernetics & Systems*, *29*(7), October, 661-688.

Rivolli, A., Garcia, L. P., Soares, C., Vanschoren, J., & de Carvalho, A. C. (2018). Towards reproducible empirical research in meta-learning. arXiv preprint arXiv:1808.10406, August, 32-52.

Smith-Miles, K. A. (2009). Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Computing Surveys (CSUR), 41(1), January, 1-25.

Suraj, Z. (2004). An introduction to rough set theory and its applications. *ICENCO, Cairo, Egypt*, 3, December, 80.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., & Bouchard, G. (2016). Complex embeddings for simple link prediction. In International conference on machine learning, June, 2071-2080.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Vanschoren, J. (2018). Meta-learning: A survey. arXiv preprint arXiv:1810.03548, October.

Watson, I., & Marir, F. (1994). Case-based reasoning: A review. *The knowledge engineering review*, 9(4), December, 327-354.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. Journal of Big data, 3(1), December, 1-40.

Yang, B., Yih, W. T., He, X., Gao, J., & Deng, L. (2014). Embedding entities and relations for learning and inference in knowledge bases. arXiv preprint arXiv:1412.6575, December.

DECISION SCIENCES INSTITUTE

Climate Change Risk Assessment for a University Campus

John Maleyeff
Boston University Metropolitan College
Email: maleyeff@bu.edu

Yujue Tan
Boston University Metropolitan College
Email: cntanyj@bu.edu

Jiaoxue Liu
Boston University Metropolitan College
Email: jxliu98@bu.edu

Maria Kristina
Boston University Metropolitan College
Email: marmaria@bu.edu

David Weidman
Boston University Metropolitan College
Email: dweidman@bu.edu

ABSTRACT

This article presents an approach to determining the impacts of climate change on a university campus by focusing on the Boston University Medical Campus. The methodology uses enterprise risk management paradigms that identify the climate-related risks and quantifies their impacts. The focus is on personnel impacts, along with business impacts separated into education and research activities. Results suggest that the highest risk ranking for the campus are associated with IT outages, staff/student access issues, and public and private transportation disruptions. The enterprise risk methodologies applied are robust and can be duplicated at other university campuses.

KEYWORDS: Enterprise risk management, Hazard vulnerability analysis, Climate change, Criticality assessment, Predictive modelling, Meta-analysis

INTRODUCTION

The City of Boston, Massachusetts is the home for 35 higher education institutions serving over 150,000 students. Boston is especially vulnerable to climate change impacts, and climate change is expected to compromise the ability of its universities to service their students. The City's Climate Vulnerability Assessment Report lists extreme heat, stormwater flooding, and coastal and riverine flooding as the major impacts of climate change in the Boston area (Climate Ready Boston, 2016, pp. 13-72). Some impacts are indirect. For example, a hurricane often leads to flooding, which can increase the potential for power outages and sewer damage. The

City also consists of an aging public infrastructure that was built without consideration of climate related risks. Its public transit system, building codes, street organization, underground heating, ventilation, and air conditioning (HVAC) equipment, low-lying emergency stations, and old water mains operate without significant backup capabilities should a major disruption occurs.

This article applies an enterprise risk management (ERM) framework to illustrate how a university campus should be evaluated based on its unique business model, stakeholder configuration, location, and physical structure. Prior work suggests that universities face many challenges, including their broad scope of activities, personnel backgrounds, and the need to customize each analysis. The methodology suggested by the authors is illustrated by an analysis of how climate change will likely impact the ability of the Boston University Medical Campus (BUMC) to serve its stakeholders (students, researchers, faculty, staff, etc.). The anticipated climate change impacts will likely affect the BUMC by compromising staff availability, research activities, students' access to classes, resource availability, and/or personnel access to food. Extreme weather-related events may also damage facilities and increase power outages. These events will have a cascading impact amplified throughout the BUMC; for example, its access to many outsourced IT services will be disrupted when power failures impact important information technologies.

BACKGROUND

ERM is a profession that covers the identification, quantification, and mitigation of risks across an organization, including operational risk, strategic risk, and financial risk (Lam, 2014, P. 10). ERM professionals seek to create organizational resiliency in the presence of risks and vulnerabilities that can impact their organization's ability to operate effectively in the short or long term. However, barriers exist in many organizations, including over-confidence, development plans that do not account for risk, and the need to customize the analysis to account for specific features of each sub-organization (Sapountzaki, 2022, P. 12-13). A hazard vulnerability analysis (HVA) is a tool that assesses risks that concerns "naturally occurring events, technological events, human-related events, and events involving hazardous materials" (Fifolt, 2016).

Researchers have addressed resiliency on university campuses using a variety of methods. Hites et al (2013) used focus groups to determine the perception of safety on a university campus. Badajoz and Caelian (2020) used a survey to show that multiple campuses across a large university need to treat the campuses differently because of their unique risk profiles. Some studies consider a subset of a university's portfolio. For example, Young et al (2023) look at the resiliency of the portfolio of academic majors. De los Reyes et al (2022) considered the resiliency associated with teaching faculty.

Recently, the COVID-19 pandemic has created an intense interest in ERM at many colleges and universities in the US (e.g., Yamey and Walensky, 2020) and abroad (e.g., Wang et al, 2020). Challenges exist in these settings because universities often operate like small cities, with their own police forces, transportation systems, and other public infrastructures (Mitroff et al, 2006). As such, they face challenges not unlike other public sector organizations, including top leadership exposure to ERM, resource limitations, and many undocumented processes (Maleyeff, 2014).

The disruptions associated with the COVID-19 pandemic should convince university administrators that anticipating climate change impacts is important, and not only because climate change may make pandemics more common (Marani et al, 2021). Stein (2023) suggests that universities need to play a socially responsible leadership role by confronting climate change. The impact of climate change on universities has not been studied extensively, although many researchers have studied how curricula need to adapt (e.g., Fahey, 2012) or how the university can play a role in public education on climate change (e.g., Hess and Maki, 2019). Some researchers have focused on carbon emissions that universities generate caused by factors such as international student mobility (Shields, 2019) and the prevalence of faculty air travel (Baer, 2023).

Boston Climate Change Projections

The BUMC is located in the south end of Boston where flooding and extreme weather are expected to increase in intensity. Its elevation is equal to sea level, and flooding potential exists from both rivers and the ocean. The adverse effects of climate change, such as increased sea-level rise, flooding, storms, and extreme heat, amplify the occurrence of hazardous events. For example, during a severe storm in 2014, many of Boston’s parking lots and public walkways experienced significant flooding. Events like these are expected to become more frequent in the future.

Douglas & Kirshen (2022) predicted more intense extreme weather events in Boston, including increases in maximum temperature (Figure 1), flood height (Figures 2 & 3), and sea-level rise (Figure 4). These predictions are based on the representative concentration pathways (RCP) index, where higher values represent more significant greenhouse gas and aerosol emissions (Meinshausen et al, 2011). Consequently, the potential for hazardous events to impact the city's infrastructure and daily life is likely to escalate, underscoring the importance of proactive planning and adaptation measures.

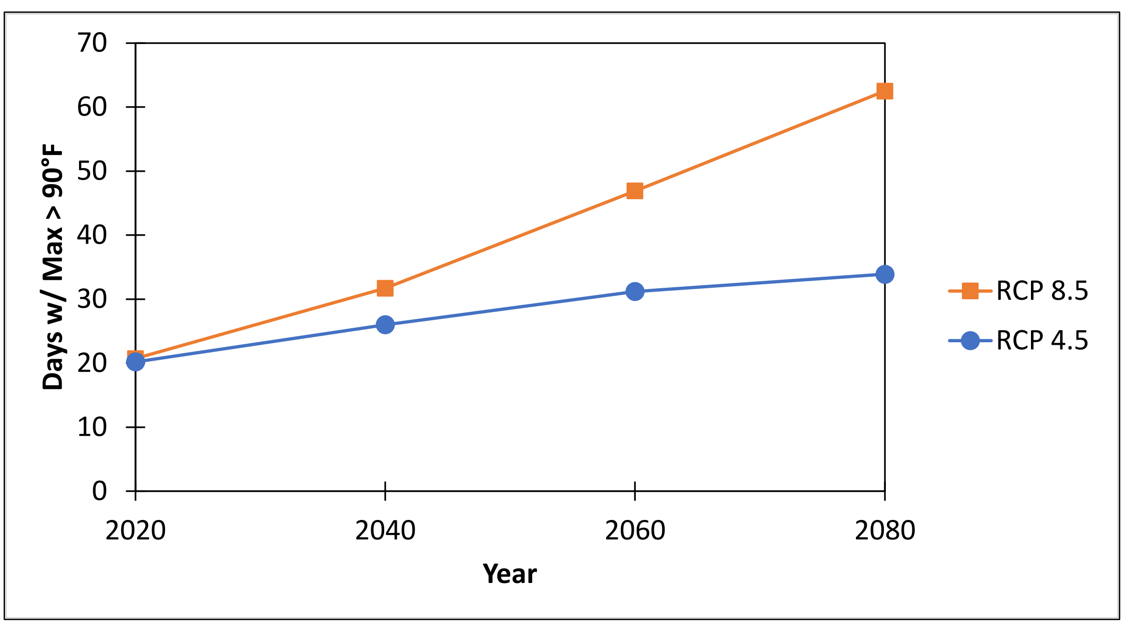


Figure 1: Projections for Number of Days w/Max Temp > 90 (°F) in Boston

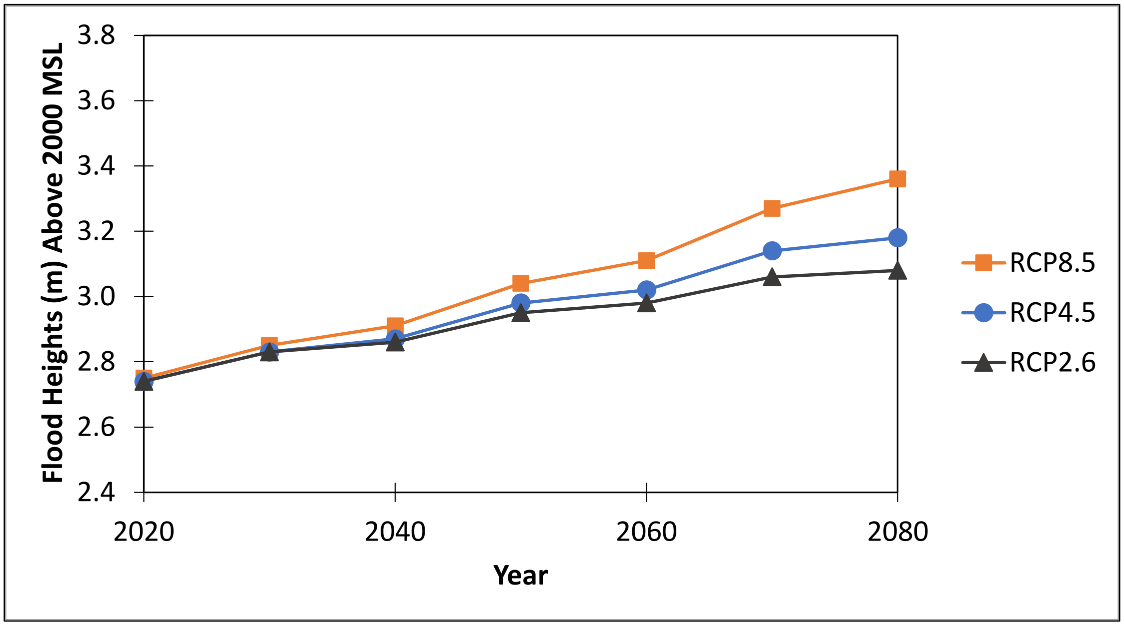


Figure 2: Projections of 10-year Winter Median Flood Heights in Boston

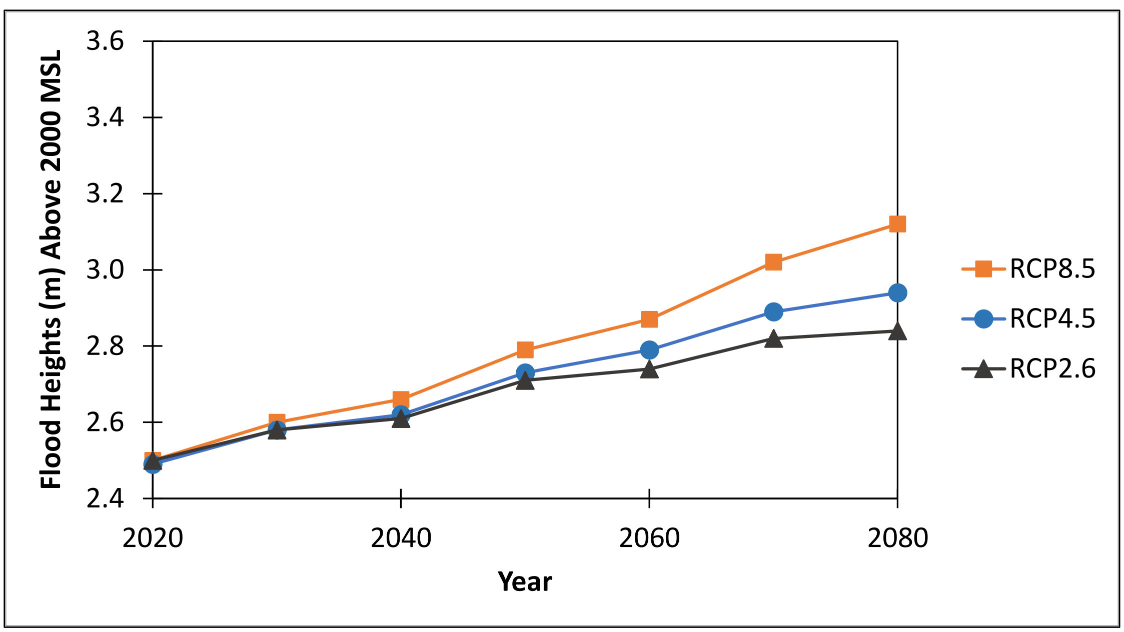


Figure 3: Projections of 10-year Summer Median Flood Heights in Boston

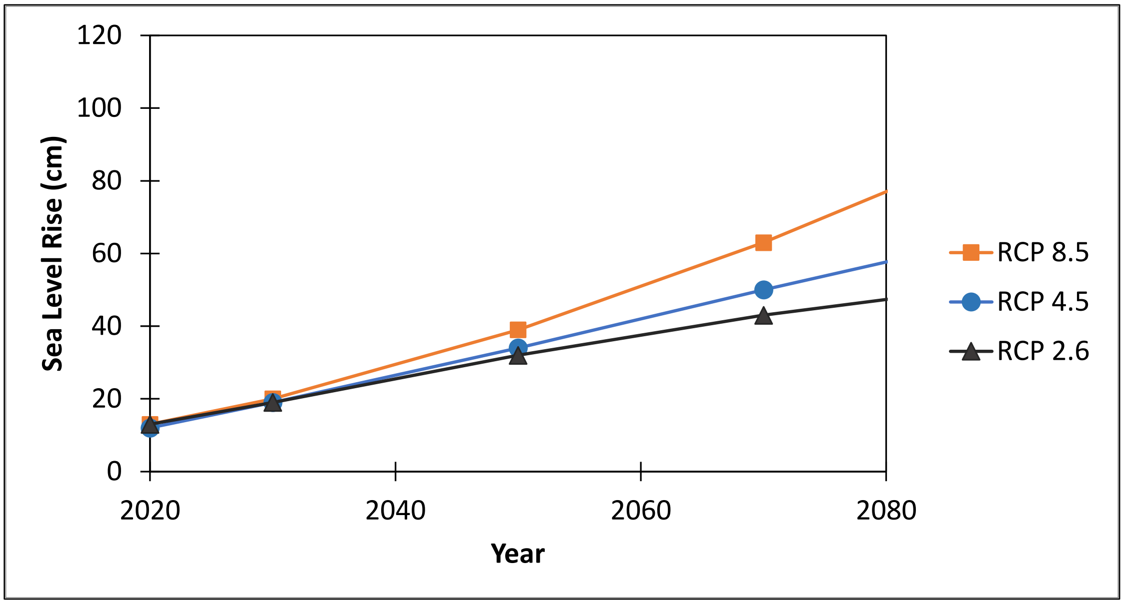


Figure 4: Projections of Median Relative Sea Level Rise for Boston Harbor

METHODOLOGY

The article's methodology consisted of an ERM assessment that incorporates meta-analysis and predictive modelling. First, the physical infrastructure of the BUMC was evaluated by touring the campus and interviewing management personnel. This evaluation resulted in a criticality analysis of the BUMC infrastructure in terms of the business activities that take place in each location. Results are documented using a scorecard as well as a display that maps the relationship among the BUMC buildings and the impacts of climate changes on the locations' ability to maintain business operations. Second, a HVA was completed based on the meta-analysis of published literature and an analysis of data relevant to climate change impacts in Boston. Input weather data were obtained from the National Oceanic and Atmospheric Administration (NOAA) that includes historical weather and extreme weather events by zip code in daily increments. Response data were also obtained from public sources.

RESULTS

The BUMC criticality analysis focused on its facilities and uncovered important risk-associated information regarding the business operations in each location (i.e., each building within which business activities take place). One important outcome of this assessment is that the BUMC facilities maintain a clear distinction between educational activities and research activities. In most cases, the research focused buildings did not include educational activities, and the classroom focused buildings did not include research activities. In only a few cases, notably where some laboratories are used for both research and as classrooms, did dual-use facilities exist.

Table 1 presents the results on the criticality analysis of each important building on the BUMC. The scale for criticality scoring was: Insignificant (Score 1), Low (Score 2), Moderate (Score 3), High (Score 4), and Essential (Score 5). Most of the criticality levels are at the extremes -

insignificant or essential, which is consistent with the distinction between the education and research activities that take place within each facility.

Table 1: Criticality Analysis

Facility Name	**Address**	**Critical Level**	
		Education	**Research**
Center for Adv Biomedical Research	700 Albany St	1	5
Biosquare Building	670 Albany St	5	3
Evans Biomedical Research Ctr – X Bldg	650 Albany St	1	5
Evans Biomedical Research Ctr – E Bldg	75 East Newton St	3	5
Dermatology- J Building	609 Albany St	1	3
School of Dental- G Building	635 Albany St	5	1
School of Medicine - M Building	75 East Concord St	1	1
School of Medicine – L Building	72 East Concord St	5	4
School of Medicine – A Building	72 East Concord St	5	1
School of Medicine – R Building	72 East Concord St	3	4
Medical Research – K Building	71 East Concord St	1	5
School of Medicine – B Building	750 Harrison Ave	2	1
School of Public Health – T Building	715 Albany St	5	3
School of Public Health – Crosstown	801 Mass Ave	1	4
Parking Garage	610 Albany St	2	2

Figure 5 shows a mapping of the climate change effects expected in Boston (more frequent and extreme flooding, storms, extreme heat, and sea level rise), and how these effects will impact the BUMC facilities. The impacts are color coded by three categories: human impacts, physical impacts, and business impacts. These categories are further subdivided as shown in the display. Colors are used to match each facility with the impacts expected to affect that facility's activities. For example, climate change will disrupt activities of the Biosquare building by limiting students' and faculty/staff access, compromising building infrastructure, and restricting medical equipment operation and supply.

The HVA shows estimates of probabilities and impacts associated with climate change effects expected in Boston's south end. It includes probability predictions for each potential disruption and impact projections for each disruption should it occur. The estimated probabilities and impacts are scored using a 1-5 scale on a relative basis. HVA accuracy is dependent on consistency in scores among the risk factors rather than an absolute quantitatively derived basis for each score.

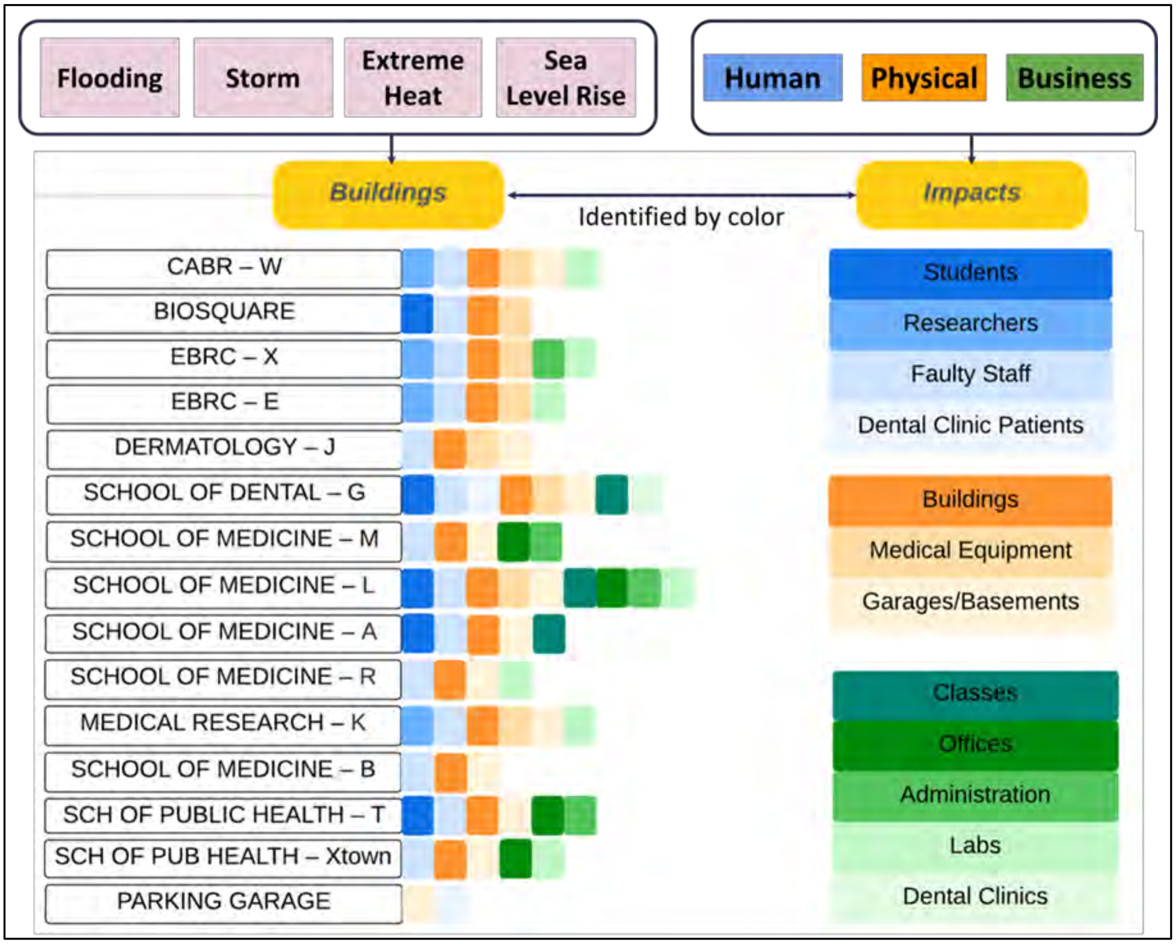


Figure 5: Display of Impacts

The scoring scale for probabilities is as follows: Insignificant (Score 1), Low (Score 2), Moderate (Score 3), High (Score 4), and Inevitable (Score 5). These scores are justified by an analysis that combines climate change projections (i.e., Figures 1-4) and estimates of specific impacts based on the meta-analysis and predictive data analysis. The impacts are separated by education and research activities, with scoring as follows: Insignificant (Score 1), Low (Score 2), Moderate (Score 3), High (Score 4), and Catastrophic (Score 5). Justification for these probability estimates and impact scores is based on the meta-analysis and data analysis (detailed in the text that follows the table). The score for risk ranking is determined by summing the impacts (education plus research) and multiplying this sum by the probability. Risk projections for the current time period, as well as 2040 and 2060, are shown in Table 2.

Justification for the scoring estimates is detailed in this section. For educational impacts, the authors considered the new remote learning enhancements at BU introduced during the COVID-19 pandemic, so that today most educational activities can be successfully conducted online. The online work option is less feasible for many research-related activities.

Table 2: HVA for the BUMC

EFFECT	YEAR	PROBABILITY	IMPACT		RISK RANKING
			EDUCATION	RESEARCH	
Power Outage	2023	1	2	3	5
	2040	2	2	3	10
	2060	3	2	3	15
IT Outage	2023	2	4	4	16
	2040	3	4	4	24
	2060	3	4	4	24
Water Failure	2023	1	1	3	4
	2040	2	1	3	8
	2060	2	1	3	8
Natural Gas Incident	2023	2	2	3	10
	2040	3	2	3	15
	2060	3	2	3	15
Extended Enterprise	2023	1	1	4	5
	2040	1	1	4	5
	2060	1	1	4	5
Last Mile Delivery	2023	2	1	4	10
	2040	3	1	4	15
	2060	3	1	4	15
Private Transit Access	2023	2	2	4	12
	2040	3	2	4	18
	2060	4	2	4	24
Public Transit Access	2023	3	3	4	21
	2040	4	3	4	28
	2060	4	3	4	28
Staff/Student Access	2023	2	2	4	12
	2040	3	2	4	18
	2060	4	2	4	24
Staff/Student Illness	2023	1	2	5	7
	2040	2	2	5	14
	2060	2	2	5	14
Air Quality	2023	1	2	3	5
	2040	2	2	3	10
	2060	2	2	3	10

Power Outage

Extreme weather can cause disruptions to power grids (Stone et al, 2023). For example, in 2012 Hurricane Sandy resulted in 15,000 outage locations and affected more than 500,000 customers in Connecticut (Caron et al, 2013). The analysis uses a quantitative approach based on response data derived from Boston's historical power outage information obtained from Eversource (the primary electricity provider in Boston) from 2017 to 2021, along with historical weather data from NOAA. It projects both the number of outages and the number of customer's affected, as shown in Equations 1 and 2 (were Y_O is the number of outages per day, Y_C is the number of customers affected per day in thousands, $X_H = 1$ when the day has high heat intensity, $X_F = 1$ when the day has severe flooding, $X_W = 1$ when the day has high wind intensity, $X_T = 1$ when the day has thunderstorms).

$$Y_O = 438 + 111X_H + 648X_F + 669X_W + 612X_T \tag{1}$$

$$Y_C = 1.6 + 4.2X_H + 2.9X_F + 3.7X_W + 3.5X_T \tag{2}$$

The HVA result is based on these results and several factors including the climate event impact to the power distribution availability, the scale of power outage, and the accessibility of alternative structures. Some research equipment, such as sterilized or hazardous equipment, requires a constant power supply and cannot be easily replicated in an online setting.

IT Outage

Although power outages can disrupt cell phones, television, home telephones and Internet services, IT infrastructures can be affected when power is not disrupted. During Hurricane Sandy, for example, 25 percent of region's FCC cell towers were not operating, Verizon's central offices had telecom equipment that was flooded, and other providers had service disruptions of varying degrees (Reuters, 2012). The impacts of IT outages can be substantial for both education and research because remote learning and work-from-home capabilities will be compromised. The HVA result is based on several factors, including the close association between power outages and IT network issues, and the varied IT infrastructure and services delivery systems including wireless, fiber, cable, satellites, and towers.

Water Failure

Floods can damage water treatment infrastructure, affecting the availability of safe and potable water for drinking (Wade et al, 2014). Ambient temperature is an important determinant of water quality because temperature compromises the ability of chlorine to kill bacteria (Hua et al, 1999). Andrade et al (2018) reported an association between floods and waterborne infections or enteric diseases via groundwater contamination. The HVA result is based on several factors, including the severity of the weather events and the effects of weather events on the city's road accessibility.

Natural Gas Incident

The aging distribution of natural gas within the City of Boston has already been shown to have substantial and widespread leaks throughout the system (Phillips, 2011). The vulnerability of the system will increase as sea levels rise, because much of the soil that supports and surrounds these pipes and joints has not been exposed to regular flooding. Soil stability will become a growing concern as flooding becomes more frequent and the structural support for the gas piping becomes less reliable. The HVA result is based on several factors, including the relationship between IT outages and power outages, the volume and magnitude of existing gas leaks throughout the Massachusetts distribution system, and the effects of future sea level rise and flooding on the city's gas infrastructure. However, natural gas is a primary heat source for only a small number of buildings on the BUMC.

Extended Enterprise

Like most enterprises, BU outsources a significant portion of its support activities and therefore can be susceptible to disruptions in their ability to provide good and services to their campuses. According to BU Environmental & Safety (2023), several prominent vendors are currently under contract with BU, providing a wide range of essential services and products, including Accutome (eye care equipment and instruments), Bayer Health (health care products and solutions), Henry Schein (surgical materials and medical equipment), and Sigma Aldrich (pharmaceutical supplies and chemicals). These and other BU suppliers are multinational corporations that maintain extensive global networks and implement contingency plans to ensure supply chain resilience in the face of potential disruptions. Although the likelihood is low, the HVA result indicates that research activities could be significantly impacted because laboratories depend on various equipment and supplies. This risk assessment is based on several factors, including the severity of the weather events and the effects of weather events on the city's road accessibility.

Last Mile Delivery

Last mile delivery (i.e., the local distribution of good to specific homes and businesses) has become an important aspect of many organizations' ability to service its customers (Joselow, 2020). These services will be disrupted when drivers' availability is compromised or when road access is limited (Gopal & de Miquel, 2017). The HVA result is based on several factors, including the severity of the weather events and the effects of weather events on the city's road accessibility, as well as the knowledge that last mile delivery is used for lab equipment (e.g., pipettes, scales, centrifuges, Bunsen burners, freezers, and hot plates).

Private Transit Access

A Fort Point Associates (2021) survey showed that about a third of the BUMC staff and students utilize private travel modes, including driving alone (23%), carpooling (3%), and biking (9%). Key driving routes to BUMC include I-93 (a primary Interstate Highway exclusive to the region), I-90 (the Massachusetts Turnpike), and the very popular Storrow and Memorial Drives. Flooding due to sea level rise can disrupt travel in Boston's artery and tunnel system, as well as sections of Storrow and Memorial Drives. The HVA result indicates that research activities are at risk because of several factors including the severity of the weather events and the effects of weather events on the City's road accessibility.

Public Transit Access

The Fort Point Associates (2021) survey also indicated that about half of the BUMC staff and students utilize public travel modes, including subway (21%), commuter rail (12%), and bus (15%), that are operated by the Massachusetts Department of Transportation (MassDOT) and the Massachusetts Bay Transportation Authority (MBTA). Climate change may present a considerable risk to the MBTA's rail rapid transit system (Martello, 2020) under the assumption that the Boston Harbor flood risk model's projections prove accurate in representing sea levels (Miller, 2019). The MBTA offers a comprehensive bus service to BUMC (MBTA, 2023), which serves as a crucial means of public transportation for students and staff traveling to the medical center. Many subway and bus routes are at risk of flooding (Martello, 2020). The HVA result is informed by the projected sea-level rise in 2040 and 2060, leading to a higher likelihood of disruptions to public transit access.

Staff/Student Access

This category covers the ability of staff and students to enter and make use of BUMC buildings and other facilities. The likelihood of this risk will increase from current levels and will impact research functions. The impact will be exacerbated due to the HVAC equipment found in the lower levels of the buildings. The HVA result is based on several factors, including the level of sea rise, the impact of flood to the campus building area, and the critical activity on each building.

Staff/Student Illness

Many researchers have studied the current and future impact of climate change on public health. These impacts can affect the BUMC by reducing the availability of workers and disrupting educational activities. Power outages can increase the incidence of food and water borne diseases (Deng et al, 2022). Rising temperatures will increase the prevalence of a heat-related illness (HRI), such as heat cramps, heat exhaustion, and heat strokes (Khan, 2019; Fuhrmann et al, 2016). High humidity also impacts the incidence of a HRI (Ortega et al, 2016). In Massachusetts, tick-borne diseases are also projected to increase (McDermott, 2022; MDPH, 2017). Romanello et al (2021) found that the Boston region can see increases in various Vibrio bacteria-related diseases such as gastroenteritis and sepsis. The HVA result considers the current low incidence of a HRI, tick-borne, and bacteria related disease in Boston so a sharp increase in their frequency will likely not significantly impact the campus. The activities of any research-focused staff or students who do contract one of these diseases will be disrupted substantially.

Air Quality

Although many adverse effects of poor air quality on the BUMC can exist, this analysis focuses on the asthma-related health impacts because the relationship of air quality to asthma has been well established (e.g., Reid et al, 2016). The HVA result is based on several factors including 8-hour ozone exceedance trends from 2012 until 2021 (MDEP, 2022), and asthma-related 911 calls trend (Reid et al, 2016). Although the chance of a large-scale asthma outbreak affecting many people simultaneously is low, disruptions to research activities may be significant because some research requires sterile conditions and others are sensitive to environmental factors, making it difficult to proceed during episodes of poor air quality.

CONCLUSIONS

This article details an ERM analysis for a university campus. It illustrates how the combination of a criticality assessment, impact mapping, and a HVA should be used to perform a comprehensive ERM analysis. The approach taken in this work makes it clear that a fundamental step in the ERM process is to separate a university's many functions into its core activities that may not always overlap within each critical facility. Although for the BUMC analysis, educational and research needed to be bifurcated, other universities or higher education institutions may be well served to consider the teaching impacts only.

Over the period covered in this work, the highest risks for the BUMC involve IT outages, staff/student access issues, and public and private transportation disruptions. The results of this project will assist BU management in further identifying and developing climate-ready resiliency strategies. Future work at BU should include the development of risk mitigation strategies while cooperating with Boston and Massachusetts agencies to coordinate mitigation policies.

ACKOWLEDGEMENTS

This work, undertaken by researchers at the Boston University Metropolitan College Decision Sciences Laboratory (DSLab), was supported by the Boston University Institute for Global Sustainability (IGS) as a Campus Climate Lab project. The authors thank the IGS for supporting this work and facilitating important interactions with relevant campus personnel. The authors also acknowledge the contributions of DSLab researchers Mengya Zhao, Pennart Klanwari, Jiayi Pan, and Xiaomeng Niu who contributed to a previous stage of this research.

REFERENCES

Andrade, L., O'Dwyer, J., O'Neill, E., & Hynds, P. (2018). Surface water flooding, groundwater contamination, and enteric disease in developed countries: A scoping review of connections and consequences. *Environmental Pollution, 236,* 540-549.

Badajoz, M., & Caelian, M.V. (2020). Vulnerability and risk assessment of a state university in the Philippines. *Philippine Social Science Journal, 2*(2), 9-22.

Baer, H.A. (2023). Grappling with climate change and the internationalization of the higher education: An eco-socialist perspective. *Journal of Studies in International Education*, 11-16.

BU Environmental & Safety. (2023). BU contracted vendor websites and phone numbers. https://www.bu.edu/ehs/ehs-topics/controlled-substances/bu-contracted-vendors/.

Caron, M.A., House, A.H., & Betkowski J.W. (2013). PURA investigates into the performance of Connecticut's electric distribution companies and gas companies in restoring service following Storm Sandy. Connecticut Public Utilities Regulatory Authority Docket 12-11-07, 64, http://www.ct.gov/pura/lib/pura/pressreleases/2012/2012march26nudraftdecision.pdf.

Climate Ready Boston (2016). Climate Vulnerability Assessment. https://www.boston.gov/sites/default/files/imce-uploads/2017-01/crb_-_focus_area_va.pdf.

De los Reyes, E., Blannin, J., Cohrssen, C., & Mahat, M. (2022). Resilience of higher education academics in the time of 21st century pandemics: a narrative review. *Journal of Higher Education Policy and Management, 44*(1), 39-56.

Deng, X., Friedman, S., Ryan, I., Zhang, W., Dong, G., Rodriguez, H., ... & Lin, S. (2022). The independent and synergistic impacts of power outages and floods on hospital admissions for multiple diseases. *Science of the Total Environment, 828*, 154305.

Douglas, E. & Kirshen, P. (2022). Climate Change Impacts and Projections for the Greater Boston Area: Findings of the Greater Boston Research Advisory Group. https://greenribboncommission.org/wp-content/uploads/2022/06/GBRAG_report_05312022@1915.pdf

Fahey, S.J. (2012). Curriculum change and climate change. *Journal of Curriculum Studies, 44*(5), 703-722.
Fifolt, M., Burrowes, J., McPherson, T., & McCormick, L.C. (2016). Strengthening emergency preparedness in higher education through hazard vulnerability analysis. *College and University, 91*(4), 61.
Fort Point Associates (2021). 2021-2031 Institutional Master Plan. https://www.bu.edu/community/files/2022/01/Boston-University-Medical-Campus-2021-2031-IMP.pdf.
Fuhrmann, C.M., Sugg, M.M., Konrad, C.E., & Waller, A. (2016). Impact of extreme heat events on emergency department visits in North Carolina (2007–2011). *Journal of community health*, *41*, 146-156.
Gopal, G., & de Miguel, A. (2017). Tackling the difficult last mile. *Industrial and Systems Engineering at Work, 49*(12), 28-33.
Hua F., West J., Barker R., & Forster C. (1999). Modelling of chlorine decay in municipal water supplies. *Water Research, 33*(12), 2735-2746.
Hess, D.J., & Maki, A. (2019). Climate change belief, sustainability education, and political values: Assessing the need for higher-education curriculum reform. *Journal of Cleaner Production, 228*, 1157-1166.
Hites, L.S., Fifolt, M., Beck, H., Su, W., Kerbawy, S., Wakelee, J., & Nassel, A. (2013). A geospatial mixed methods approach to assessing campus safety. *Evaluation Review, 37*(5), 347-369.
Joselow, M. (2020), Delivery vehicles increasingly choke cities with pollution. *Scientific American*, E&E News, https://www.eenews.net/.
Khan, A.A. (2019). Heat related illnesses: Review of an ongoing challenge. *Saudi Medical Journal*, *40*(12), 1195-1201.
Lam, J. (2014). *Enterprise Risk Management: From incentives to controls (2/E).* John Wiley & Sons, New York.
Maleyeff, J. (2014) Sustaining public sector Lean Six Sigma: Perspectives from North America, *Management and Organizational Studies, 1*(2), 92-99.
Marani, M., Katul, G.G., Pan, W.K., & Parolari, A.J. (2021). Intensity and frequency of extreme novel epidemics. *Proceedings of the National Academy of Sciences - PNAS, 118*(35), 1.
Martello, M.V. (2020). *Resilience of rapid transit networks in the context of climate change.* Doctoral Dissertation, Massachusetts Institute of Technology, Boston, Massachusetts.
MBTA (2023). Full System Map, Massachusetts Bay Transportation Authority, https://cdn.mbta.com/sites/default/files/2023-03/2023-3-13-system-map.pdf
MDEP (2022). Massachusetts 2021 Air Quality Report. Massachusetts Department of Environmental Protection.
MDPH (2017). Tick-borne disease prevention. Massachusetts Department of Public Health, https://www.mass.gov/info-details/tick-borne-disease-prevention.
McDermott, A. (2022). Climate change hastens disease spread across the globe. *Proceedings of the National Academy of Sciences - PNAS, 119*(7), e2200481119.
Meinshausen, M., Smith, S.J., Calvin, K. et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300*. Climatic Change 109*, 213.
Miller, S.J. (2019). *MassDOT/MBTA Actions contained in the SHMCAP and Coastal Flood Risk Modelling.* Presentation at the MassDOT Transportation Innovation Conference, Worcester, MA. https://www.umasstransportationcenter.org/Document.asp?DocID=694.
Mitroff, I.I., Diamond, M.A., & Alpaslan, M.C. (2006). How prepared are America's colleges and universities for major crises? *Change: the magazine of Higher Education, 38*(1), 61-67.

Ortega, A., Sarmiento, J.M., Ly, D., Nuño, M., Mukherjee, D., Black, K.L., & Patil, C.G. (2016). Multiple resections and survival of recurrent glioblastoma patients in the temozolomide era. *Journal of Clinical Neuroscience, 24*, 105-111.

Phillips, N. (2011). BU Today. *Fueling Global Warming, Not Homes*. Boston University. https://www.reuters.com/article/storm-sandy-telecommunications/update-3-hurricane-sandy-disrupts-northeast-us-telecom-networks-idUSL1E8LUEFU20121030.

Reid, M., Gunn, J., Shah, S., Donovan, M., Eggo, R., Babin, S., ... & Burkom, H. (2016) Cross-disciplinary consultancy to enhance prediction of asthma exacerbation risk in Boston. *Online Journal of Public Health Informatics 8*(3), e199-e199.

Reuters. (2012). *UPDATE 3-Hurricane Sandy disrupts Northeast US telecom networks.* https://www.reuters.com/article/storm-sandy-telecommunications/update-3-hurricane-sandy-disrupts-northeast-us-telecom-networks-idUSL1E8LUEFU20121030.

Romanello, M., McGushin, A., Di Napoli, C., Drummond, P., Hughes, N., Jamart, L., ... & Hamilton, I. (2021). The 2021 report of the Lancet Countdown on health and climate change: code red for a healthy future. *The Lancet, 398*(10311), 1619-1662.

Sapountzaki, K. (2022). Risk mitigation, vulnerability management and resilience under disasters. *Sustainability, 14*(6), 3589.

Shields, R. (2019). The sustainability of international higher education: Student mobility and global climate change. *Journal of Cleaner Production, 217*, 594-602.

Stein, S. (2023). Universities confronting climate change: Beyond sustainable development and solutionism. *Higher Education, 2023*.

Stone Jr, B., Gronlund, C.J., Mallen, E., Hondula, D., O'Neill, M.S., Rajput, M., ... & Georgescu, M. (2023). How blackouts during heat waves amplify mortality and morbidity risk. *Environmental Science & Technology*, Article ASAP.

Wade, T.J., Lin, C.J., Jagai, J.S., & Hilborn, E.D. (2014). Flooding and emergency room visits for gastrointestinal illness in Massachusetts: a case-crossover study. *PLoS One*, *9*(10), e110474.

Wang, C., Cheng, Z., Yue, X.G., & McAleer, M. (2020). Risk Management of COVID-19 by universities in China. *Journal of Risk and Financial Management, 13*(2), 36.

Yamey, G., & Walensky, R.P. (2020). Covid-19: Re-opening universities is high risk. *BMJ (Online), 370*, m3365-m3365.

Young, M., Pinheiro, R., & Avramovic, A. (2023). Unpacking resilience in higher education: Investigating twenty-first century shifts in universities' academic cores. *Higher Education (Online).*

DECISION SCIENCES INSTITUTE

Closet-Building versus Minimalism: Selling Fewer, Better Products to Fashion Sensitive Customers

Monire Jalili
Bentley University
Email: mjalili@bentley.edu

Michael S. Pangburn
University of Oregon
Email: pangburn@uoregon.edu

Alireza Yazdani
California State Polytechnic University
Email: ayazdani@cpp.edu

ABSTRACT

Fashion sellers are sometimes critiqued for selling products with low durability, resulting in waste. Blame is also directed at consumers, who purchase new fashions despite having accumulated a closet full of prior fashions. The ``slow fashion'' movement encourages sellers to produce more durable products, thus supporting less frequent purchases by consumers. We analyze a seller facing a market of consumers who differ in their sensitivity to fashion, in a setting where fashion changes over time. Using an infinite-time model and considering strategic consumer behavior, including their ability to accumulate a ``closet'' of varieties over time, we analyze the seller's profit-maximizing price and product-durability decisions. We initially assume a static price but later analyze a scenario with dynamic pricing. When analyzing a heterogeneous consumer market, we initially allow customers to vary (distributed uniformly) in their sensitivity to fashion. Subsequently, we explore alternative distributions for consumers' fashion sensitivity and the correlation between their fashion sensitivities and product valuations. Using this framework, we show how the seller's optimal price and durability decisions yield distinct shopping segments, which we refer to as the minimalist versus closet-building behaviors. We find that if the degree of fashion uncertainty is moderate, the seller's optimal choice of product durability will support the coexistence of both behaviors. As the variety uncertainty expands, if the seller's costs are sufficiently low, it will support a throwaway culture via disposable products. Otherwise, given high costs, the seller optimally targets a slow-fashion type outcome, with consumers targeting reuse (with durability) rather than variety. Our findings shed light on consumers' optimal purchasing behaviors in relation to both market parameters and the firm's pricing and durability decisions, and we show these findings remain robust relative to modeling perturbations.

KEYWORDS: Product durability, pricing, retail, fashion, market segmentation

DECISION SCIENCES INSTITUTE

Co-creation in New Product Development: Collaborating with a Shared Supplier in the Presence of a Competitor

Abhishek Roy
Temple University, Fox School of Business
Email: abhishek.roy@temple.edu

ABSTRACT

Firms are increasingly involving their suppliers in co-creating new products and services through collaborative alliances. In many industries, such suppliers are shared by competing manufacturers, who may benefit from forming alliances and co-creating new products with their shared suppliers. Using a game theoretic model, we analyze the strategic interactions in such a collaborative alliance when competing buyers decide to co-create a common component with the supplier. We demonstrate that competing buyers may sometimes benefit from collaborating jointly with the shared supplier in co-creating common components, instead of shirking away from the collaboration or letting the supplier be the sole developer of the component.

KEYWORDS: Collaboration, Co-production, Co-creation, Shared Supplier, Competition, Game Theory, Analytical Modeling

DECISION SCIENCES INSTITUTE

Cookie & Card Game: Understanding Machine Learning Algorithms Using Humanizing Pedagogy

Sudipendra Nath Roy
St. Cloud State University
snroy1818@gmail.com

ABSTRACT

We live in a world where after buying a laptop from Amazon, we see advertisements for laptop bags in our Facebook feeds. This necessitates us to understand business analytics and machine learning concepts to be successful in the business environment. While the knowledge of machine learning is imperative in today's world, students often avoid technical and quantitative courses due to apprehension. We suggested a classroom activity using cards and cookies to teach three basic machine learning algorithms (unsupervised, supervised, and reinforcement learning) as a part of an introductory analytics course and evaluated the activity with students of an applied statistics course.

KEYWORDS: Machine learning, Business Analytics, Active learning

DECISION SCIENCES INSTITUTE

COVID-19 Uncertainty and Corporate Risk-taking: International Evidence

Md Ismail Haidar
University of Texas Rio Grande Valley
E-mail: mdismail.haidar01@utrgv.ed

Md Showaib Rahman Sarker
University of Texas Rio Grande Valley
E-mail: mdshowaibrahman.sarker01@utrgv.edu

ABSTRACT

Using data from 187,925 firm-quarters observations and 25,940 firms in 71 countries from January 2020 to June 2022, this study shows that COVID-19 uncertainty increases corporate risk-taking, especially for firms in common law, developed, and less financially constrained countries. The study also suggests that cash holdings and agency issues mediate the effect of COVID-19 uncertainty on corporate default risk, implying that firms may take more risks to conserve cash and reduce agency conflicts amid the pandemic. The findings highlight the importance of prudent risk management for firms facing high COVID-19 uncertainty.

KEYWORDS: COVID-19 uncertainty, Corporate risk-taking, Agency problem

DECISION SCIENCES INSTITUTE

Diabetes Mellitus Trend among Adults in Texas Counties, from 1999-2009: A Spatial Analysis of Events

Abah Theresa[a] & Oppong Joseph[b]
[a]Assistant Professor, Department of Gerontology, California State University, Sacramento, CA. 95819.Email: t.abah@csus.edu*
[b]Professor, Department of Geography and the Environment, University of North Texas, Denton. TX. 76203. Email: joseph.oppong@unt.edu
*Corresponding author

ABSTRACT

Diabetes mellitus is a chronic disease of public health concern globally; it is estimated to be the seventh leading cause of death by 2030. About 7 million people are diagnosed of the disease yearly, with an estimated 1.5 million deaths in 2019. While Texas is ranked among the top 10 states with the highest diabetes disease prevalence rates in the U.S. This study examined association between ethnicity and other variables and how they produce diabetes disease, guided by the Disease Ecology Framework. It provides recommendations on utilizing informed decision-making as a strategy in prevent diabetes, considering the evolving healthcare landscape.

KEYWORDS: Diabetes disease, Metabolic syndrome, Spatial analysis, Built-environment, Place vulnerability

INTRODUCTION

Diabetes mellitus is a chronic disease of public health concern globally; it is estimated to be the seventh leading cause of death by 2030. About 7 million people are diagnosed of the disease yearly, with more cases recorded among adult populations. In 2019, an estimated 1.5 million deaths were as a result of complications from diabetes mellitus, along with other top ten causes of death such as, ischemic heart disease, stroke, lower respiratory infections, kidney diseases, lung cancers and Alzheimer disease among others. In the US, about 37.3 million people have diabetes disease, and 1 in 3 people are at the risk of dying from the diabetes disease yearly. While Texas is ranked among the top 10 states with the highest diabetes disease prevalence rates in the country with 12%, compared to the national average of 9.8%, the obesity rate of 36.1% mirrors the national average of 36%. Within the state of Texas, a mortality rate of 70 per 100,000 births was observed as the state level, while there was 29 per 100,000 births within the north-central counties, and counties bordering New Mexico, Mexico, the Panhandle and Gulf of Mexico. The Disease Ecology Framework was used to provide some insights into some of the risk factors associated with diabetes disease in Texas counties. Race / Ethnicity, Obesity, Physical Inactivity, Genetics, Nutrition and social Economic status (SES) are some risk factors associated with the disease. The association between Race / Ethnicity and Age-Adjusted Death Rate (AADR) of diabetes disease for the period of evaluation -1990 to 2009 show strong correlation between the disease and population subgroup, in addition to all other variables.

Diabetes mellitus is a condition where the pancreas no longer produces insulin, or when cells stop responding to the insulin that is produced. It accounts for about 1.5 million deaths globally (World Health Organization, WHO Global report in 2020). This increasing trend in global diabetes disease mortality rate is attributed to increase in obesity, behavioral health patterns, globalization, poor nutrition, increase in Body Mass Index (BMI), a correlation of genetics / epigenetic factors and Severe Acute Respiratory Syndrome Coronavirus 2, SARS-CoV-2 (Glovaci et al., 2019; Oguntibeju, 2019; Kumar et al., 2020). There are two types of the disease Type 1- which is a total lack of insulin, and Type 2 - little or ineffective use of the body's insulin. They both have different mortality rates, with type 2 diabetes having the highest mortality rates. The Diabetes Association of America estimated the economic cost of treating diabetes disease at $245 billion in 2012 and $966 in 2021. In addition, the CDC report of 2021 on diabetes disease suggest that about 37.3 million Americans have diabetes (about 11.3% of U.S. population), with 15.9 million making up older adults 65 years and older (CDC.gov).

Other studies have shown that, adults are at a higher risk of developing, and dying from the disease especially those with, a family history of the disease, people who are overweight, those with special health problems like high blood pressure and people who are physically inactive (Statista, 2022). Furthermore, the Texas Department of State Health Services, DSHS report of 2022 suggest that the prevalence of the disease in the US differs by region. The southern regions near Mexico border have mortality rates of 13.6%, while the south east region had 10.9%, with the lowest mortality rate at the North East region with 10.3%. (Texas, DSHS Report, 2022). Despite the concentration of diabetes cases within the southern region - the US –Mexico border region, also noted to have lower SES, the age adjusted death rate, AADR of the disease within, the north and north east region of the state, the north-central, and north eastern regions, are reported to have higher diabetes mortality rates when adjusted by population density (approximately 28 deaths per 100,000 births to zero deaths in counties in the southern region of Texas), Texas, DSHS Report (2022).

This study examined the geography of diabetes mortality rates in Texas counties with the goal of understanding the trend of diabetes disease and factors that influence it, focusing on few variables such as, race and ethnicity, environment and SES. Other risk factors such as physical inactivity, smoking habit, quality of life, length of life and age interactions are used to deepen our understanding of the disease. This is the first part of a series of studies that would explore diabetes disease in Texas counties starting from 1999 to 2009. The research questions guiding this study are as follows: (1) Does diabetes mortality rate increase by race / ethnicity? (2) Does the rate of diabetes increase by place of residence? (3) Does an individual's SES level increase their risk of dying from diabetes?

LITERATURE REVIEW

Diabetes and Race / Ethnicity /SES

The CDC (National Health and Nutrition Examination Survey (NHANES) report of 2020 suggest race / ethnicity continues as well as cultural differences within the same race show an association with diabetes. With a national prevalence rate of 9.8%, the age adjusted, AA diabetes disease prevalence rates for Hispanics was almost twice as high as non-Hispanic Whites, (21.2% versus 12%), while the AA prevalence diabetes disease rates for non-Hispanic Black was 18.8% and non-Hispanic Asian 18.1%. In addition, Hispanic males have a higher

prevalence rate of the disease than females (23.5% versus 19.2%), supporting the difference in the disease trend even within ethnic populations. The geographic distribution of diabetes diagnosis among U.S. states with high population of Hispanics compared in 2021 was, 9.7% for California, Texas 12%, Illinois 9%, New York 9.8%, Florida 10.5%, and Puerto Rico 13.1%. The percentages and rates are age-adjusted to reflect the US census standard population.
Extant studies, report there are significant differences in diabetes prevalence rates between the main ethnic groups in the U.S., (Rodríguez & Campbell, 2017; Hannis, 2016; Harrison, 2011 and Foy, 2006). Diabetes prevalence rates by race was 11.7% for Blacks, 9.6% for Hispanics and 4.8% among non – Hispanic Whites. Also, the World Health Organization, WHO Diabetes Report in 2015 sums up the deaths due to diabetes as 1.5 million worldwide, with differences in prevalence rates by geographic regions. For example, South East Asia and Western Pacific regions recorded almost half of the world's total cases of 422 million adult mortality, with the rest of the world accounting for the remaining numbers (WHO, Report 2015). Mokdad and colleagues in 2000 examined diabetes trend in the US using the Behavioral Risk Factor Surveillance System (BFRSS) data set from 1990 to 1998 to examine non- institutionalized adults. About 33% prevalence rate of diabetes was observed within the review period among males and females when controlled for educational level and race (Mokdad et. al., 2001). Overall, there was a difference of about 3% higher prevalence rates of diabetes disease by race, age, gender, and education level among population subgroups.

Diabetes and Behavioral Pattern

Evert and colleagues in 2014 examined the impact of behavioral modification and diabetes. A correlation between smoking and diabetes in adults was examined for the associated risk of diabetes while controlling for a combination of cardiovascular disease and insulin resistance. Males were categorized into smokers and non-smokers within the period. Findings showed that 25% of smokers had were reported to have developed diabetes within 5-year period while 14% of non-smokers developed diabetes within the same period. However, the nutrition and behavior pattern for the group were not considered (Davison et al., 2013; Evert et. al., 2014). Another study by Ford and colleagues examined long-term patterns of weight change and diabetes risk using NHANES data. The findings show that an increase in weight within the range of 13kg to 20kg or more resulted in 27% higher risk for diabetes incidence, even for steady weight gain of over 5kg. Suggesting that, an increase in BMI results in increased rates of diabetes, although the data to establish this relationship between weight gain and the genetic composition of the participants was not evaluated (Ford et al., 2002). Both BMI and weight gain are major risk factors of diabetes, a 1kilogram increase in weight resulted in approximately 9% prevalence rate. In addition, another CDC 2001 report examined the influence of risky behaviors and diabetes prevalence. Using the BFRSS, a substantial increase in the prevalence of obesity was recorded between 1999 and 2001 with increase in obesity among adults. Variations were also observed for cigarette smoking, binge drinking, and heavy drinking

Diabetes and SES

In 2004, a CDC report examined the association between SES disadvantage and diabetes, specifically how low education attainment produce diabetes. A large variation in diabetes prevalence was observed by state and age group. Further, a diabetes report by the WHO and National Institutes of Health show similar association between diabetes and low income globally. Diabetes prevalence in high income countries was estimated at 3% - 8%, while low income countries was between 6% and 14%. This situation is reversed now, as at 2021, the upper middle-income economies have 8.2% prevalent rate, lower middle-income economies, 7.1%, and low-income economies 6.7% prevalent rates. Adding that, 1 in 3 US adults are obese, due to inability to afford nutritious diets, which further increases the risk of diabetes (NIH, 2021; WHO, 2015, and CDC, 2004).

Maty and colleagues in 2005, examined the relationship between diabetes prevalence and socio-economic status. Using reported information of education, income and occupation, it was observed that these variables had an association with increased risk of the disease. The highest predictor was low education- groups with less than 12% of education had 50% risk of diabetes compared with those with more education. The other variables like environment and behavior had minimal association with diabetes disease compared with low education (Maty et. al., 2005).

Diabetes and the Built Environment / Physical Inactivity / Place vulnerability

The geography of a place has been shown through various studies to influence health outcomes due to the interactions of neighborhood, living environment, built environment / physical activity (Bower, 2012; Coxe & McCullough, 2015; Deitz, 2016; Tiraphat, 2012). Tiraphat in a study in 2012, examined the influence of environmental conditions on physical activity engagement among different population sub groups in different geographic regions, and suggest that, there's significant correlation between physical activity engagement and the built environment. Non-Hispanic Whites benefited more from the built environment than other races because they had access to parks, bike lanes and walk ways that are located within their environments compared with other ethnic groups. People from White ethnic groups were more likely to be physically active, while minority groups like Asians, African Americans and Hispanic populations. In addition, minority populations living in counties with high income bracket were less active than those living in low income bracket counties. This surprising finding suggest that, people's ethnicity also determines their likelihood to be physically active, and as such, influence the probability of developing diabetes.

In 2012, Bower in their study on the influence of neighborhood composition and food store availability, observed that, racial segregation has been linked to the poor health of African Americans. They hypothesized that more segregated black neighborhoods - characterized by limited nutritious food stores, would have poor participation in physical activity. Which is consequently tied to obesity, causing a disparity among races. Similarly, Stoddard and colleagues in 2013, observed an association between the built environment and risk of type 2 diabetes. After moderating for availability of spaces to exercise to stay physically active, they observed that adults who were deprived of a safe space were less likely to lose weight / BMI loss, and vice visa. This resulted in increased risk of type 2 diabetes among these groups, with a BMI change (16.1 % gain in weight and 17% loss in weight), Stoddard (2013). The presence of a park and exercise area were associated with meeting physical activity requirement and vice visa.

THEORETICAL DEVELOPMENT/MODEL

The theory guiding the interpretation of the results from this study is the Disease Ecology Framework. The supporters of this theory argue that disease ecology varies by geographic regions, which is further influenced by multifactorial causes within, and outside an individual's environment (Oppong & Harold, 2009). For example, an individual's genetics plays a vital role in their overall health outcome, however, the influence of environment, behavior and social status predisposes them to diseases, and at different intensity.

Disease Ecology Framework – An Interdisciplinary Theory

An interdisciplinary effort which includes the following disciplines, biology, ecology, geography, epidemiology, veterinary, sciences, etc., providing understanding on the influence of environmental factors on health and wellbeing. The framework tries to explain why diseases persist despite the huge resources and efforts deployed by all stakeholders to eradicate them in society.

Disease Ecology Framework: Key Assumptions and Tenets

The key tenets of the concept are that, (1) there's a need to link community health in disease ecology for better programmatic solutions and (2), the need to examine how socio-economic status acts as a driver to good or poor health outcomes among population subgroups.

METHODS

The data used for this study was obtained from several sources. Diabetes data was downloaded from Vital Records control, VITALWEB of the Texas Department of health – this is an end-to-end electronic records management software which provides online access to health records for all Texans (www.https:/vitalweb.vitalrecordscontrol.com). Diabetes disease is classified as ICD-9 code 124. The records used include Age–Adjusted Death Rates, AADR by county, as recorded from 1999 through to 2009. The data for race, ethnicity and gender were obtained from the 2010 American Community Survey (www.census.gov). Lastly, the data for other control variables like, adult-smoking history, adult-physical inactivity and adult obesity records were downloaded from the Texas County health rankings, 2016 (www.countyhealthrankings.org). The 2010 population census data set was utilized for this study. Race/ethnicity and gender data sets were calculated in percentages, and collected for different time-period. These statistics provided ample guide for reporting the results. Unlike previous studies that focused their evaluation on estimating diabetes disease prevalence by states, we focused the evaluation at county level, to provide deeper insights into disease trend to prevent late diagnoses of the disease (Congdon & Lloyd, 2021; Fisher et. al., 2015).
For the result analysis, statistical statistics software, SPSS provided insights on the statistical significance of the predictive variables for diabetes disease mortality. Simple correlation between diabetes disease mortality and risks from all other independent variables, such as, obesity, physical inactivity, race/ethnicity, socio-economic status, SES and access to health care were also analyzed. To determine how the identified risk factors, predict the prevalence of diabetes disease mortality in each county, multiple regression analysis was used to evaluate the results, (Field, 2009).

RESULTS

Spatial distribution of Diabetes among Adults

The mortality rate of diabetes across the counties varied considerably with the southern- most part of the state having a high concentration of Hispanics population, and with the highest rates of diabetes. The AADR as illustrated in Figures 1- 4, shows the spatial distribution of diabetes in all the counties, the mortality rates and a change map for mortality and prevalence rates between 1999 and 2009. The diabetes prevalence and the mortality rates by race/ ethnicity represented in Figures (5-6), the dominant ethnic group within the northern counties are non-Hispanic White population. dominate this region. In addition, portions of the mid-west and western regions, predominantly non- Hispanic White populations had high clusters of counties with high diabetes rates (44.8-70.4%). See Figure 1. A few counties in the eastern region, with a high concentration of Black population, including the south-eastern regions had high diabetes mortality rates (49.1-82% and 37.7-49%). See Figure 2. It is imperative to note however that, analysis from the county level data only provides a crude result of the actual mortality rates due to the heterogeneous nature of the counties in terms of size, population and geographic distribution of the population. A more accurate diabetes mortality spatial analysis requires finer geographic units having homogeneous County patterns.

The number of counties with AADR higher than the state average of 28.6 per 100,000 diabetes cases were 177 out of 254. The five counties with the highest mortality rates include; Frio (70.4), Duval (62.2), Schleicher (57.7), Shackelford (54.2), Dawson (53.8), and Willacy (52.5) counties had the highest mortality rates. These Counties are located within the North-Central region, North East and Central regions of the state. See Figure (2 & 3). Despite the large area of the Southern border, the recorded AADR of diabetes was highest in the north central region. See Figure 4. More than half of the state had diabetes mortality rates above the state average, between 30 to 70 per 100,000 populations. Also, the counties with the lowest diabetes mortality rates include; King (0), Loving (0), Oldham (7.3), Borden (9), and Baylor (10.4), all located within the north and north western region of the state. See Figure (1-4).

Diabetes changes over time

The trend of diabetes mortality within the 10-year review period showed a steady rise from 1999 to 2002, with a decline from 2003 to 2009. See Figure 4. The highest AADR recorded in 2003 was 32 per 100,000, while the lowest AADR was in 2009 at 23.3 per 100,000. A number of factors could be responsible for the decline in mortality rates, ranging from improved access to care, better program planning, focus on the thematic areas of the disease, etc. The change map highlights this change, Fig 4. It shows the change in AADR among the north central, north east and central regions of the states.

Diabetes and Race / Ethnicity

We hypothesized that Race is a predictor of diabetes mortality. Counties with high concentration of Hispanic population will have higher diabetes mortality, while predominantly White counties will have lower diabetes mortality. The diabetes mortality rate (Fig 5) shows a concentration of cases around the southern border, a predominantly Hispanic population, however our statistics did not support that hypothesis. Our observation showed that counties with the highest mortality rates were within the northeastern, north central and northwestern regions. In addition, the diabetes mortality rate was statistically significant, with correlation coefficients for (Hispanics at 0.337, Blacks at 0.117 and Whites at -0.435). There was negative correlation, between diabetes mortality and White population. The result shows a correlation between diabetes mortality rates by race/ethnicity, this was evident among Hispanic and Black population sub groups. Further analysis of the data, Fig (1- 4) show that the counties with the lowest disease cases were recorded in Oldham and Borden with 7.3 and 9.0 mortality rates per 100,000 population. These Counties are located within the extreme north and North West regions of the state; which is a predominantly White settlement. Four counties within the central part of the state had high diabetes mortality rates ranging from 41.6 to 54.2 per 100,000 population, they include Shackelford, Coleman, Callahan and Runnels.

The correlation coefficients for Hispanics as earlier stated was 0.337, Blacks was 0.117 and Whites negative 0.435 respectively. The bar charts equally show that the highest concentration of AADR for diabetes was among the Black ethnic minority groups (50 per 100,000 births). Hispanics and non-Hispanic Whites had 43 and 21 deaths per 100,000 population, (See figures 6&7). This study supports other findings about the high mortality rate of diabetes by ethnic group (Benavides-Vaello, & Brown, 2016; Fisher-Hoch et. al., 2015). A number of factors could be responsible for the health outcomes of people from this ethnic group including, genetic, socio economic status, or/and access to good health care.

Figure 1: Map of Texas showing Diabetes AADR from 1999 to 1009

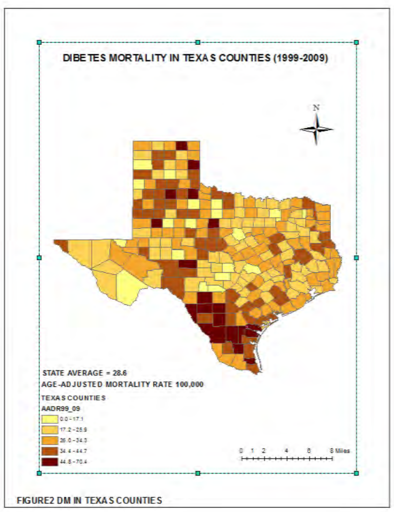


Figure 2: Diabetes mortality in Texas Counties (1999 – 2004)

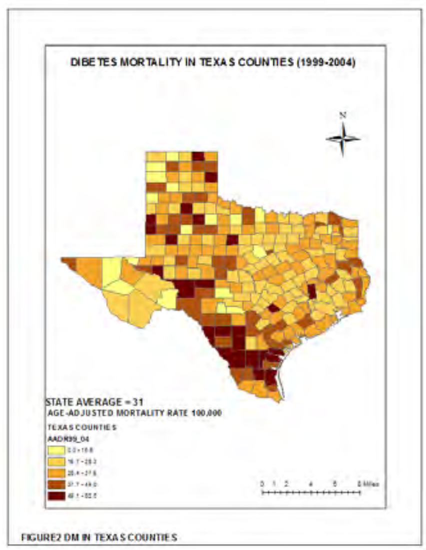


Figure 3: Diabetes mortality in Texas Counties (2005 – 2009)

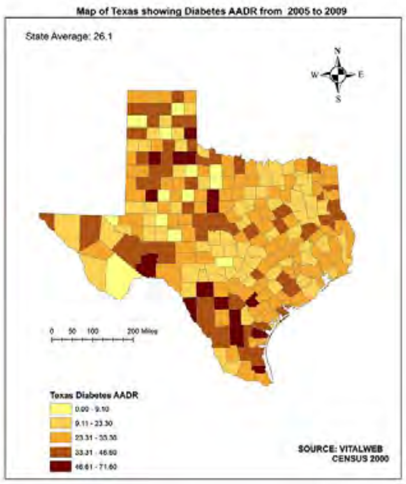


Figure 4: Diabetes AADR change map (2005 – 2009)

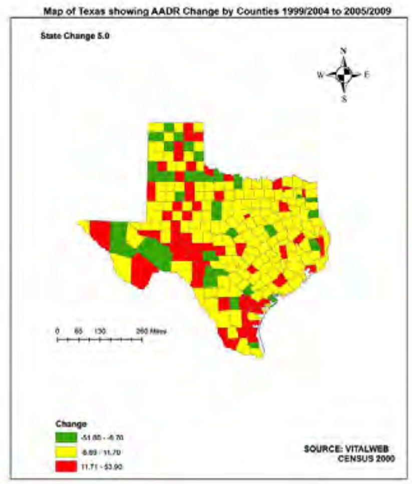


Figure 5A: Race, Ethnicity and Diabetes

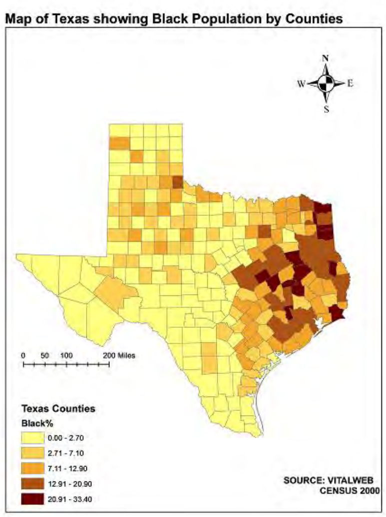


Figure 5A: Race, Ethnicity and Diabetes

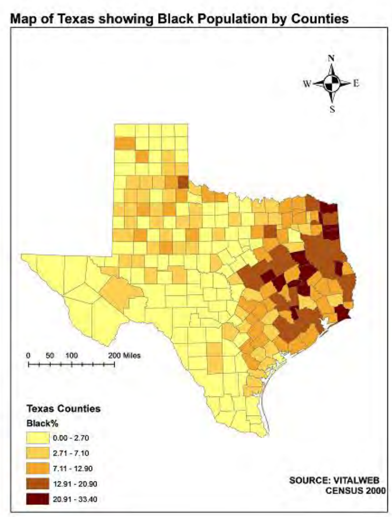


Figure 6A: Map of Texas showing White population

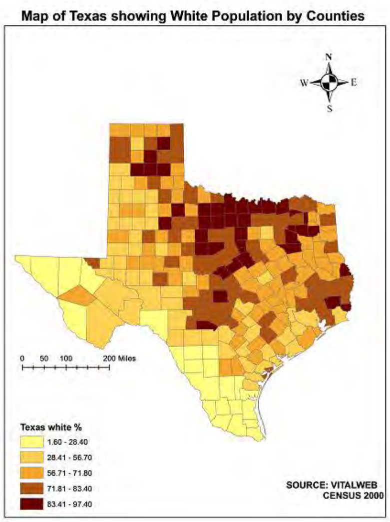


***Figure 6B:** Map of Texas showing diabetes mortality rates by counties AADR*

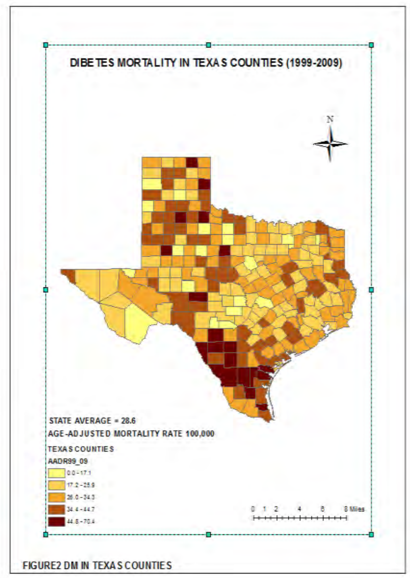


Figure 7A: Diabetes Disease AADR by years

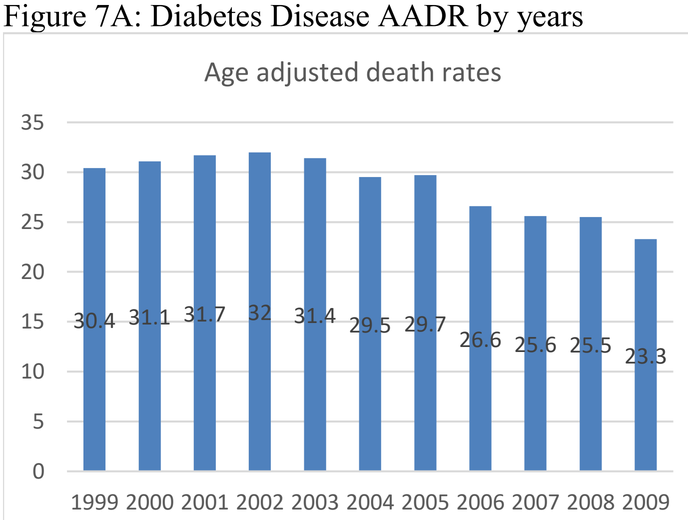


Figure 7B: Diabetes Disease AADR by Race/Ethnicity

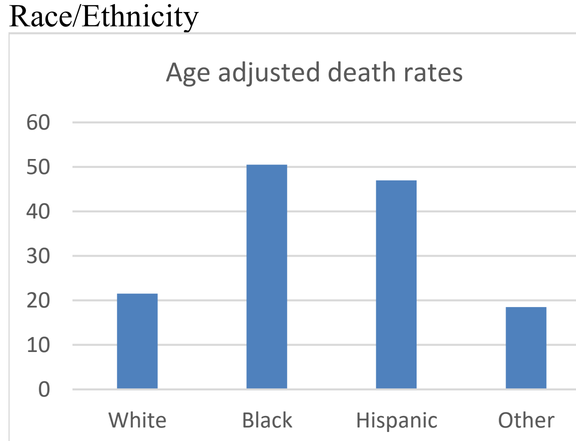


Table 1: Descriptive statistics for the outcome variable, Diabetes Disease Prevalence, among Texas Counties

Correlation Analysis Output

Variable	Pearson Correlation Coefficient	P value
Adult Obesity	0.140*	0.025
Adult Physical Inactivity	0.024	0.706
Adult Smoker	0.331*	0.000
Quality of Life	0.444*	0.000
Length of Life	0.177*	0.005
Physical Environment	0.002	0.973
SEF (Socioeconomic Factors)	0.328*	0.000
Percent Black	0.117	0.016
Percent Hispanic	0.337*	0.000
Percent White	-0.435*	0.000
Correlation is significant at 0.05 levels (2-tailed)		

Statistical Analysis

Table 1 shows the result of analysis for six variables that increase the risk of diabetes, including adult obesity, adult physical inactivity, smoking (Independent Variables, IV) and race/ethnicity (percent Hispanic, Black, White and Other races) observed in Texas Counties against the dependent variable. The dependent variables were regressed on the IVs, giving the following results.

Behavior Pattern, SES and Diabetes

Socio-economic status, SES measured by socio-economic factor, SEF computed variable which comprised (population and housing problems). We hypothesized that SES is a predictor of diabetes mortality. Counties with low SEFs will have corresponding higher rates of mortality, vise visa. A statistically significant relationship was observed between diabetes mortality rate and SEF. A correlation of 0.328 suggests that diabetes mortality rate increases as the SEF reduces. In addition, low-income areas within the Hispanic and Black ethnic population groups, which had had high diabetes cases (47 per 100,000) and Black (50 per 100,000) cases compared to Whites (20 per 100,000). See Figure 7.

Similarly, controlling for adult smokers, obesity and physical inactivity, there was a significant correlation with diabetes mortality rates (Table 1). This supports findings in the literature (Deitz, 2016 and Lin, 2015), that behavior pattern influence diabetes mortality rates, and it is higher by race/ethnic group. There was statistical significance for correlation between the disease two independent variables, with 0.140 for obesity and 0.331 for adult smokers and a p-value (at $p<.05$). Counties with higher numbers of adult smokers will consequently have higher diabetes mortality rates than those with fewer smokers. In addition, counties with more physically active people will have lower diabetes mortality rates. The literature reviewed suggested that White ethnic group are more active than other ethnic groups, this finding further confirms this submission. Despite the large population of non-Hispanic White populations as compared to other groups, they still recorded the lowest diabetes mortality rates across the counties. Furthermore, existing literature suggest that, non – Hispanic Whites have better access to health care services because of wider coverage of health insurance while other ethnic groups have lower health insurance coverage as a result poorer health care (Stoddard et. al., 2013; Benavides-Vaello & Brown, 2016).

Built Environment, Place Vulnerability and Diabetes

Studies have suggested that the built-up environment -which includes natural amenities such as parks and gardens, walk-ways, good street orientation, permeability, good landscaping etc., (Stoddard et. al., 2013; Benavides-Vaello & Brown, 2016), are required to maintain good health and prevent illness. It has been determined that non – Hispanic Whites benefit more from the built-up environment, while other groups are less likely to benefit due to cultural orientation, awareness of the benefits, cost of exercising against competing tasks, etc. The statistical analysis shows a weak correlation between the built environment and diabetes mortality. Length of life and quality of life are used as a measure of place vulnerability. The statistically significant correlation of 0.444 with a p-value, $p<.05$ for quality of life supports this assumption. Similarly, a strong correlation of 0.177 with a p-value, $p<.05$ was recorded for Length of life. See Table 1. Studies also suggest that, other ethnic minority groups do not have as much access to health care services, consequently leading to higher vulnerability and higher disease burden and mortality rates for people from low SES background. This further explains why the black ethnic group has the highest diabetes mortality rate despite their lower population compared to the non – Hispanic Whites and Hispanics.

Multiple regression output

The multiple regression analysis was utilized to determine how the outlined risk factors (Table 1), predict diabetes mortality rates. It was also used to explain the association between the independent variables.

Table 2: Model summary from SPSS output

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.551[a]	0.304	0.275	8.67760

a. Predictors: (Constant), race/ethnicity, behavior pattern / SES and built environment / place vulnerability

Table 3: Coefficients from multiple regression analysis

Table 3 shows the statistical output of the various predicting variables of diabetes disease in all Texas counties for the period in review.

Coefficients[a]

MODEL	BETA COEFFICIENT
Constant	10.341
Quality of Life	0.203
SES	-0.131
Physical Environment	0.026
Length of Life	0.203
Smokers	0.012
Obesity	-0.004
Physical Inactivity	0.014
White	0.550
Black	0.373
Hispanics	1.074

a. Dependent variable: Diabetes Disease

Table 3 above sheds more insights on diabetes disease trend in Texas counties. The R square value of .304 was obtained from the multiple regression analysis performed, suggesting a 30.4% chance of the occurrence of diabetes mortality by the predictor variables (race/ethnicity, behavior pattern / SES and built environment / place vulnerability. The difference in the total sum of 69.6% is associated with other factors (example access to care, nutrition, genetics and epigenetic factors) that can explain diabetes mortality, but that is not covered in this study. Furthermore, it can be deduced that a unit increase in the Hispanic ethnic group population will result in the increase of the diabetes AADR, same applies for Blacks and Whites, however, the diabetes AADR will be more among Hispanic groups.

DISCUSSION

This study sought to examine diabetes disease trend in Texas counties, we observed that SES, lack of physical activity, obesity, environment and race / ethnicity are risk markers of the disease and increase diabetes disease mortality in Texas along geographic regions. With the exemptions of physical environment which showed a weak correlation with the disease, and percent White, which had an inverse association, the other risk factors and markers had direct relationship with diabetes mortality rates. These associations provide insight into the importance of physical activity, better lifestyle practices and healthy diets on the part of the individual among other social interventions available. A close look at the trend of the disease within the same geographical location provides further information about the impact of race / ethnicity mix and diabetes mortality rates in Texas. The diabetes mortality AADR change map shows that despite the downward trend in deaths, the same counties continue to record high mortality rates. Prevention efforts and improvements in early diagnosis, treatment, and care must be responsible for the reduction of deaths between 2005 – 2009. Other beneficial interventions could be improved medical care, more health personnel knowledgeable about the resources available to communities, including funding support.

Proponents of the Disease Ecology Framework theory (Johnson, Escobar, & Zambrana-Torrelio, 2019; Oppong & Harold, 2009) argue that diseases are endemic in different geographic regions, while other regions do not have as much episodes. This study supports that assumption, it goes further to show that the multifactorial causes of increased diabetes disease cases among low income population groups does not explain the low mortality rates for individuals from the southern counties in Texas, although they experience the highest death rates of the disease.
The limitation of this study is remains with the use of proxy data to explain some of the variables, and conducting the study at the micro level (County level data) rather than within a fine scale (zip code level) – which would have provided a more heterogeneous data. With more homogeneous data and smaller scale, the better the representation of events and findings.
In addition, the data sets used for the analysis of deaths for the period under review were analyzed using a more recent data set for the hypothesized variables even though the deaths happened over 20 years ago. The difference in time makes it difficult to make firm conclusions about the findings. On its merit however, the evidence provided suggest that diabetes disease is still an endemic disease that needs close monitoring, and further programmatic support to reduce the disease.

CONCLUSION

The study sought out to determine if diabetes mortality rate increases with behavior change, race / ethnicity, physical environment, SES or/and an association of all these predictors. It was observed that all the listed variables / predictors of the disease either had a positive or negative correlation with diabetes mortality rate. It is appropriate to add that within the study period of ten years (from 1999 to 2002), AADR increased steadily, thereafter we recorded a decline of the disease for about six years, from 2003 to 2009. In all, race/ethnicity and SES continue to be a driving factor for the high mortality rates. If this trend continues, individuals with the disease stand the risk of developing metabolic syndrome, further complicating the disease (James et al., 2020). Furthermore, the unavailability of nutritious food sources could be driving the concentration of diabetes within the southern borders– noted with high cases of the disease. This lends credence to the assumption that the geography of a place affecting the health outcome of the people (Serban, 2015 and Nash, 2010). This study adds to the knowledge of diabetes disease risks, prevention and metabolic disease issues by focusing on the influence of race / ethnicity, geography and other predictive factors that produce the disease. It concludes by making the following recommendations.

Decision-making by all stakeholders in diabetes disease prevention and management plays a crucial role. With the evolving healthcare landscape and technological advancements, this global health issue could be reversed in this age of transition. Some interventions to adopt include the use of Artificial Intelligence (AI) and Predictive Analytics; Patient Empowerment; and Eliminating food desserts among others. These strategies can be deployed with the use of technology to help transform the current diabetes disease surveillance and management landscape. Also, program managers can provide relevant and responsive decisions that can transform patient outcome using available data with prompt attention focused on problematic areas.

With the evolving healthcare landscape and technological advancements, future studies on this topic should utilize more recent data in the decision-making process, and examine the risk of the disease by gender - for more improved health outcomes. Researchers should do this by investigating among ethnic groups and across groups for a more efficient allocation of resources in preventing the disease.

REFERENCES

Benavides-Vaello, S., & Brown, S. A. (2016). Sociocultural construction of food ways in low-income Mexican American women with diabetes: A qualitative study. *Journal of Clinical Nursing*, 25(15-16), 2367-2377. doi://dx.doi.org/10.1111/jocn.13291

Bower Joffe, K. M. (2012). *Race, place, and obesity among women in the united states: The impact of neighborhood racial composition and food store availability on racial disparities in obesity* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (1288382440).

Behavioral Risk Factor Surveillance System (BFRSS) data (199 https://www.cdc.gov/brfss/index.html

Center for Disease Control and Prevention (CDC). (2004). Prevalence of Diabetes among Hispanics — Selected Areas, 1998–2002. MMWR 53, 941-944.

Center for Disease Control and Prevention (CDC). (2008). 2005 BRFSS Codebook. http://www.cdc.gov/brfss/technical info data/survey data/2005.htm

Congdon, P., & Lloyd, P. (2010). Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. *Journal of Data Science*, *8*(2), 235-252.

Coxe, L. M., Lennertz, K., & McCullough, S. (2015). Changing prevalence of diabetes for Texas nursing home residents, 1999-2009. *Journal of Applied Gerontology*: The Official Journal of the Southern Gerontological Society, 34(8), 938-957. doi://dx.doi.org/10.1177/0733464813512180.

Davison, K. K., Jurkowski, J. M., & Lawson, H. A. (2013). Reframing family-centered obesity prevention using the Family Ecological Model. *Public Health Nutrition*, *16*(10), 1861–1869. https://doi.org/10.1017/S1368980012004533

Deitz, S. L. (2016). *A spatial analysis of the relationship between obesity and the built environment in southern Illinois* (M.S.). Available from ProQuest Dissertations & Theses Global. (1803939442).

DHS Report 2020, accessed at: https://www.dshs.texas.gov/about-dshs

Evert, A. B., Boucher, J. L., Cypress, M., Dunbar, S. A., Franz, M. J., Mayer-Davis, E. J., Neumiller, J. J., Nwankwo, R., Verdi, C. L., Urbanski, P., & Yancy, W. S. (2014). Nutrition therapy recommendations for the management of adults with diabetes. *Diabetes Care*, *37*(SUPPL.1), 120–143. https://doi.org/10.2337/dc14-S120

Fisher-Hoch, S. P., Vatcheva, K. P., Rahbar, M. H., & McCormick, J. B. (2015). Undiagnosed diabetes and pre-diabetes in health disparities. *PloS One*, 10(7), e0133135. doi://dx.doi.org/10.1371/journal.pone.0133135

Ford, E. S., Giles, W. H., & Dietz, W. H. (2002). Prevalence of the metabolic syndrome among US adults: Findings from the Third National Health and Nutrition Examination Survey.

Journal of the American Medical Association, *287*(3), 356–359. https://doi.org/10.1001/jama.287.3.356

Foy, C. G., Foley, K. L., D'Agostino, R. B., Goff, D. C., Mayer-Davis, E., & Wagenknecht, L. E. (2006). Physical activity, insulin sensitivity, and hypertension among US adults: findings from the Insulin Resistance Atherosclerosis Study. *American journal of epidemiology*, 163(10), 921-928.

Glovaci, D., Fan, W., & Wong, N. D. (2019). Epidemiology of Diabetes Mellitus and Cardiovascular Disease. *Current Cardiology Reports*, *21*(4). https://doi.org/10.1007/s11886-019-1107-y

Hanis, C. L., Redline, S., Cade, B. E., Bell, G. I., Cox, N. J., Below, J. E., Aguilar, D. (2016). Beyond type 2 diabetes, obesity and hypertension: An axis including sleep apnea, left ventricular hypertrophy, endothelial dysfunction, and aortic stiffness among Mexican Americans in Starr county, Texas. Cardiovascular Diabetology, 15, 86. doi://dx.doi.org/10.1186/s12933-016-0405-6.

Harrison, W. W., Bearse, M. A., Ng, J. S., Jewell, N. P., Barez, S., Burger, D., ... & Adams, A. J. (2011). Multifocal electroretinograms predict onset of diabetic retinopathy in adult patients with diabetes. *Investigative ophthalmology & visual science*, *52*(2), 772-777.

Healthy Vision - (May 2022,). MMWR. Morbidity and Mortality Weekly Report Retrieved from http://search.proquest.com/docview/1684295606

James, K. (2010). *Arsenic exposure in drinking water and the associated risk with coronary heart disease and type II diabetes mellitus* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (757675836).

Johnson, E. E., Escobar, L. E., & Zambrana-Torrelio, C. (2019). An ecological framework for modeling the geography of disease transmission. *Trends in ecology & evolution*, *34*(7), 655-668.

Kumar, A., Arora, A., Sharma, P., Anikhindi, S. A., Bansal, N., Singla, V., Khare, S., & Srivastava, A. (2020). Is diabetes mellitus associated with mortality and severity of COVID-19? A meta-analysis. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, *14*(4), 535–545. https://doi.org/10.1016/j.dsx.2020.04.044

Maty, S., Everson-Rose, S., Haan, M., Raghunathan, T. and Kaplan. G. (2005). Education, income, occupation, and the 34-year incidence (1965–99) of Type 2 diabetes in the Alameda County Study. *International Journal of Epidemiology* 34, 1282-3

Mokdad, A., Ford, E., Bowman, B., Nelson, D., Engelgau, M., Vinicor, F. and Marks, J. (2001) Diabetes trends in the U.S: 1990-1998. Diabetes Care 24, 1278-1283. Mukhtar, Q., Murphy, D. and Mitchell, P. (2003). Use of data from the behavioral risk factor surveillance system optional diabetes module by states. *J. Public. Health Manag*ement. Pract. Suppl, S52-5.

Nash, A. C. (2010). *Methodological focus and application of exploratory dietary patterns in epidemiological research* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (758939378).

National Health and Nutrition Examination Survey (NHANES accessed at: https://www.cdc.gov/nchs/nhanes/index.htm

Oguntibeju, O. O. (2019). Type 2 Diabetes Mellitus, oxidative stress. *International Journal of Physiology Pathophysiology and Pharmacology*, *11*(3), 45–63.

Oppong, J. R., & Harold, A. (2009). Disease, ecology, and environment. *A companion to health and medical geography*, 81-95.

Rodríguez, J. E., & Campbell, K. M. (2017). Racial and ethnic disparities in prevalence and care of patients with type 2 diabetes. *Clinical Diabetes*, *35*(1), 66–70. https://doi.org/10.2337/cd15-0048

Serban, R. E. (2015a). *Racial/ethnic disparities in access to primary care: A spatial approach to determining the effects of residential segregation, socioeconomic status and health factors on access to primary care in U.S. adults* Available from Dissertations & Theses Europe Full Text: Business. Retrieved from http://search.proquest.com/docview/1719276301

Serban, R. E. (2015b). *Racial/ethnic disparities in access to primary care: A spatial approach to determining the effects of residential segregation, socioeconomic status and health factors on access to primary care in U.S. adults* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (1719276301).

Smyth, F. (2005). Medical geography: therapeutic places, spaces and networks. *Progress in human geography*, 29(4), 488-495.

Statista (2020 Report). Accessed at: https://www.statista.com/

Stoddard, P. J., Laraia, B. A., Warton, E. M., Moffet, H. H., Adler, N. E., Schillinger, D., & Karter, A. J. (2013). Neighborhood deprivation and change in BMI among adults with type 2 diabetes: The Diabetes Study of Northern California (DISTANCE). *Diabetes Care*, *36*(5), 1200-1208.

Texas County health rankings, 2016 (www.countyhealthrankings.org).

Tiraphat, S. (2012). *Multilevel analysis of physical activity among us adults across US census regions: The role of environmental contexts* (Ph.D.). Available from ProQuest Dissertations & Theses Global. (1140117070).

WHO, W. H. O. (2015). *World Report on Aging and Health*.

DECISION SCIENCES INSTITUTE

Discretion in Automated Supermarket Replenishment: Censorship Bias and Self-inflicted Stockouts

Bengü Nur Özdemir
IE Business School
Email: bengu.ozdemir@student.ie.edu

Antti Tenhiälä
IE Business School
Email: antti.tenhiala@ie.edu

ABSTRACT

We study a paradox in which decision-makers order less than proposals of an automatic store replenishment system after a stockout. We argue that censorship bias explains this curious ordering behavior and that it has negative performance implications. We compare the effect of censorship bias to the effect of anchoring bias to understand the relative importance of censorship bias in practice. To suggest policies, we collect more data to test the idea of blocking deviations when censorship bias is suspected. With this additional analysis, we show that by blocking downward deviations susceptible to censorship bias, retail managers can reduce self-inflicted stockouts.

KEYWORDS: Individual biases, Retail inventory, Human judgment, Decision support systems, Econometrics

DECISION SCIENCES INSTITUTE

Eco-efficiency Assessment of the OECD Countries
by DEA Models

Umit Saglam
East Tennessee State University
Email: saglam@etsu.edu

ABSTRACT

This study quantitatively evaluates the eco-efficiency of the OECD countries using five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable (real GDP), and one undesirable output variable (CO_2 emissions). The input-oriented CCR, BCC, and SBM models evaluate the states' technical and pure technical efficiency. The results indicate that only eight eco-efficient countries, Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, are designated as CRS (constant returns to scale), indicating that they operated at the most efficient level of productive efficiency in 2018. On the other hand, IRS (increasing returns to scale) is assigned to only two countries, Finland and Lithuania, implying that they might considerably increase their production by expanding their input variables. Twenty-eight countries are classed as DRS (decreasing returns to scale), suggesting that their production level improves less than the proportion growth in their inputs. Finally, Luxembourg is benchmarked by 30 other countries due to the optimum input and output variables combination. The findings of this study will enable policymakers to understand the latest relative eco-efficiency assessments of the OECD countries.

KEYWORDS: Data Envelopment Analysis (DEA), Eco-efficiency, Slack-Based Measure (SBM), Undesirable Output

INTRODUCTION

Global warming is the long-term increase in average temperatures across the Earth's surface, atmosphere, and oceans. Fourier (1827) initially identified this phenomenon, and Arrhenius (1897) developed an early model linking ground temperature to carbon dioxide concentration. Over the past five decades, the discernible and measurable impacts of global warming on human beings have become more apparent. According to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), the average global surface temperature has risen by 1°C since 1850, primarily due to a relentless surge in greenhouse gas concentrations.

The previous IPCC report indicates that if we stabilize greenhouse gas concentrations, the global average surface warming would still amount to approximately 0.2°C per decade. Even this seemingly modest increase could have severe consequences for ecosystems worldwide. As projected in the report's scenarios, a 2°C rise in temperatures would endanger around 25% of plant and animal species. Additionally, crop yields would decline, resulting in food scarcity in numerous regions. The prevalence of malnutrition would lead to health problems for millions of people, and the incidence of deaths, diseases, and injuries would escalate due to extreme weather events. Climate change is also causing a significant rise in sea levels, putting Asian, African, and Caribbean islands at risk from storm surges, inundation, and erosion. Some islands may even face the threat of disappearance altogether. In summary, if global warming is not curbed, climate change will precipitate many severe problems affecting various aspects of life.

Not surprisingly, global warming and climate change became the most critical environmental and political issues between countries. First, in 1997, 192 countries signed the Kyoto Protocol, which delimits the production of GHG emissions to fight global warming and climate change. However, it was not a complete success because the world's three major GHG polluters, China, the United States, and India, which account for more than half of global GHG emissions, did not establish binding objectives. As a result, 196 parties signed the Paris Agreement in 2015 to limit the increase in global average temperature to less than 1.5 degrees Celsius above pre-industrial levels. According to the International Energy Agency (IEA) data for 2021, OECD member countries accounted for approximately 27% of global energy-related GHG emissions. OECD countries' energy and emission efficiency are critical to accomplishing this target.

In this study, linear programming problems are modeled to quantitatively evaluate the eco-efficiency of OECD countries using five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO2 emissions). The input-oriented CCR (Charnes, Cooper, and Rhodes (1978)) and BCC (Banker, Charnes, and Cooper (1984)) models are implemented to evaluate the OECD countries' technical efficiency and pure technical efficiency. Besides, scale efficiency and returns to scale (RTS) data are computed for the OECD countries. Also, input-oriented SBM (Slack-based measure) efficiency scores are calculated to incorporate undesirable output variables. Lastly, the findings of this study will enable policymakers to understand the latest relative eco-efficiency assessments of the OECD countries.

The remainder of this paper is organized as follows: Section 2 provides a brief literature review of DEA-related articles on the eco-efficiency topic. Section 3 presents an overview of the DEA framework for the input-oriented CCR, BCC, and SBM models. Section 4 describes the data in detail by selecting input and output variables. Section 5 reports the DEA results. Finally, Section 6 provides a summary and some concluding remarks.

LITERATURE REVIEW

Data Envelopment Analysis (DEA) is a very popular management tool for evaluating and improving the efficiency of both manufacturing and service operations. It was developed by Charnes et al. (1978).

This research assesses eco-efficiency in the 38 OECD countries utilizing five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO_2 emissions). Capital stock and labor force are two critical components in all stages of production. The level of aggregate output may be measured using (real) GDP (real GDP). Energy consumption is an essential input variable in the manufacturing process and should be accounted for in the model to assess energy efficiency. However, this study incorporated primary energy consumption, including fossil fuels (petroleum, natural gas, and coal), nuclear energy, and renewable energy sources. Finally, because carbon dioxide (CO_2) emissions negatively result from manufacturing activities, they are critical for environmental efficiency and long-term economic growth. Table 1 summarizes the DEA applications in detail on this topic, considering these input and output variables. Please see Demiral and Sağlam (2021) for a thorough review of these studies.

Table 1: Summary of DEA applications in the literature

Reference	Measure	DMUs	Period	Methods
Zhou and Ang (2008)	Energy efficiency	21 OECD countries	1997-2001	DEA
Zhou et al. (2010)	Total-factor emission efficiency	Top 18 CO_2 emitters	1995-2004	MPI
Yeh et al. (2010)	Energy efficiency	China and Taiwan	2002-2007	DEA
Choi et al. (2012)	Energy efficiency	China's 30 Provinces	2001-2010	SBM
Wang et al. (2013)	Energy and environmental efficiency	China's 30 Provinces	2006-2010	DEA-window
Apergis et al. (2015)	Energy efficiency	OECD countries	1985-2011	SBM, MCMC-GLMM
Iftikhar et al. (2016)	Energy and emission efficiency	26 major economies	2013-2014	SBM
Moutinho et al. (2017)	Environmental efficiency	26 European countries	2001-2012	DEA, QR
Moutinho et al. (2018)	Eco-efficiency	16 Latin American countries	1994-2013	MPI
Park et al. (2018)	Environmental and emission efficiency	US's 50 states	2004-2012	SBM
Wang et al. (2020)	Eco-efficiency	17 European countries	2013-2017	SBM
Demiral and Sağlam (2021)	Eco-efficiency, Eco-productivity	The US's 50 states	2018	DEA, SBM
Demiral and Sağlam (2023)	Eco-efficiency, Eco-productivity	The US's 50 states	1997-2018	DEA, SBM, MPI

This study quantitatively evaluates eco-efficiency in the 38 OECD countries utilizing five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO2 emissions). The input-oriented CCR and BCC models are implemented to evaluate the states' technical efficiency, pure technical efficiency, and scale efficiency. In addition, SBM also evaluates the eco-efficiency of the OECD countries by indirectly incorporating undesirable variables into the model. Also, the benchmark countries are identified for each inefficient OECD country from the dual linear programming problems. Lastly, the findings of this study will enable policymakers to understand the latest relative eco-efficiency assessments of the OECD countries.

MODEL DEVELOPMENT

Data Envelopment Analysis (DEA)

DEA is a non-parametric, multi-factor relative efficiency measure for evaluating and improving the effectiveness of both manufacturing and service operations. Charnes et al. (1978) introduced the DEA framework to calculate the relative efficiency score by dividing the weighted sum of outputs and a weighted sum of inputs to obtain decision-making units (DMU). Equation 1 formulates the scenario where we have N number of maximized DMUs, which are achieved by s number of output and m number of input variables:

$$\begin{aligned} \textit{Max.} \quad & E_k = \frac{\sum_r^s u_r y_{rk}}{\sum_i^m v_i x_{ik}} \frac{\sum_r^s u_r y_{rj}}{\sum_i^m v_i x_{ij}} \leq 1, \forall j \\ \textit{s.t.} \quad & u_r, v_i \geq 0, \forall i, r. \end{aligned} \tag{1}$$

where E_e is the maximized efficiency score, which belongs to the e^{th} DMU.

The Input-Oriented CCR Model

The input-oriented model's objective is to minimize input variable(s) while keeping the current level of output fixed. The input-oriented CCR model can be formulated as a linear programming problem under the CRS assumption. The relative technical efficiency score of the k^{th} DMU (θ_k) with n number of DMUs, s number of output, and m number of input variables, can be modeled as follows:

$$\begin{aligned} \textit{Min.} \quad & \theta_k - \varepsilon\left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right) \\ \textit{s.t.} \quad & \theta_k x_{ik} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_{ik}^- = 0 \qquad i = 1, \dots, m \\ & y_{rk} - \sum_{j=1}^{n} y_{rj}\lambda_j - s_{rk}^+ = 0 \qquad r = 1, \dots, s \\ & \lambda_j, s_i^-, s_r^+ \geq 0; \qquad j = 1, \dots, n; \quad r = 1, \dots, s; \quad i = 1, \dots, m \end{aligned} \tag{2}$$

where s_r^+ and s_i^- represent non-negative slack variables for output and input constraints, respectively. x_{ij} represents the amount of ith input variable that j^{th} DMU consumes; y_{rj} represents the amount of r^{th} output variable that j^{th} DMU produces, and lastly, λ_j represents structural variables. Also, ε is a non-negative infinitesimal number for keeping the coefficients of input and output variables positive.

The Input-Oriented BCC Model

The input-oriented BCC model can be formulated by adding a convexity constraint to Equation 2. The relative pure efficiency score of the kth DMU (ξ_k) of BCC model under variable returns to scale (VRS) assumption can be formulated as Equation 3:

$$
\begin{aligned}
\textit{Min.}\quad & \varphi_k - \varepsilon\left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right) \\
\textit{s.t.}\quad & \varphi_k x_{ik} - \sum_{j=1}^{n} x_{ij}\lambda_j - s_{ik}^- = 0 \qquad i = 1, \dots, m \\
& y_{rk} - \sum_{j=1}^{n} y_{rj}\lambda_j - s_{rk}^+ = 0 \qquad r = 1, \dots, s \\
& \sum_{j=1}^{n} \lambda_j = 1 \\
& \lambda_j, s_i^-, s_r^+ \geq 0; \qquad j = 1, \dots, n; \quad r = 1, \dots, s; \quad i = 1, \dots, m
\end{aligned}
\tag{3}
$$

The input-oriented CCR and BCC models are adopted from Charnes et al. (1978), Bankers et al. (1984), and Sağlam (2017a, 2017b, 2017c, 2018a, 2018b, 2019, 2022).

The Slack-Based Measure (SBM)

The Slack-Based Measure (SBM) is a methodology used in Data Envelopment Analysis (DEA) to assess the efficiency of decision-making units (DMUs). Introduced by Tone (2001), SBM is a non-parametric approach incorporating input and/or output slacks directly into the efficiency measurement process. The SBM methodology is widely used in various fields, including economics, management, and environmental studies. It allows researchers and practitioners to identify inefficient DMUs, determine potential areas for improvement, and make informed decisions regarding resource allocation and productivity enhancement. There are three orientations of SBM: non-oriented, input-oriented, and output-oriented. The non-oriented SBM assesses overall efficiency without distinguishing between input and output efficiencies. In contrast, the input-oriented SBM measures the efficiency of DMUs in converting inputs into outputs, while the output-oriented SBM focuses on the efficiency of DMUs in generating outputs from given inputs. Input-oriented SBM-DEA efficiency model under the CRS assumption can be formulated as follows:

$$
\begin{aligned}
\textit{Min.}\quad & 1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_i^-}{x_{ik}} \\
\textit{s.t.}\quad & x_{ik} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- \qquad i = 1,2, \dots, m;
\end{aligned}
\tag{4}
$$

$$y^{g}_{r_1k} = \sum_{j=1}^{n} y^{g}_{r_1j}\lambda_j - s^{g}_{r_1} \qquad r_1 = 1,2,\dots,q_1;$$
$$y^{b}_{r_2k} = \sum_{j=1}^{n} y^{b}_{r_2j}\lambda_j + s^{b}_{r_2} \qquad r_2 = 1,2,\dots,q_2;$$
$$\lambda_j, s_i^-, s^{g}_{r_1}, s^{b}_{r_2} \geq 0;$$
$$j \in R^n, i \in R^m, r_1 \in R^{q_1}, r_2 \in R^{q_2}$$

DATA DESCRIPTION

This study quantitatively evaluates eco-efficiency in the 38 OECD countries utilizing five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO2 emissions). Capital stock and labor force are two fundamental components in all manufacturing phases. The aggregate output level can be assessed by (real) gross domestic product (real GDP). Energy consumption is an essential input variable in the manufacturing process and should be considered in the model to evaluate energy efficiency. Hence, this study incorporated primary energy consumption, including fossil fuels (petroleum, natural gas, and coal), nuclear energy, and renewable energy sources. Finally, carbon dioxide (CO_2) emissions are an unfavorable byproduct of manufacturing operations, making them essential for environmental efficiency and long-term economic development. Table 2 shows the Pearson correlations between pre-determined input and output variables in 2018 for the OECD countries.

Table 2: Pearson Correlations between input and output variables (p-values are given in parentheses)

Pearson Correlations	GFCF	LF	FFC	NPC	REC	GDP	CE
Gross fixed capital formation (GFCF)	1.000	0.972 (0.000)	0.983 (0.000)	0.890 (0.000)	0.930 (0.000)	0.996 (0.000)	0.985 (0.000)
Labor Force (LF)		1.000	0.937 (0.000)	0.817 (0.000)	0.875 (0.000)	0.966 (0.000)	0.940 (0.000)
Fossil Fuel Consumption (FFC)			1.000	0.893 (0.000)	0.956 (0.000)	0.987 (0.000)	1.000 (0.000)
Nuclear Power Consumption (NPC)				1.000	0.876 (0.000)	0.896 (0.000)	0.888 (0.000)
Renewable Energy Consumption (REC)					1.000	0.942 (0.000)	0.954 (0.000)
Gross Domestic Product (GDP)						1.000	0.987 (0.000)
CO_2 Emission (CE)							1.000

Input Variables

In this study, we consider three input variables for the DEA models: (1) gross fixed capital formation, (2) labor force, (3) fossil fuel consumption, (4) nuclear power consumption, and (5) renewable energy consumption. Table 3 presents the related data for these input variables.

Gross fixed capital formation (GFCF):

There is a robust correlation between gross fixed capital formation and real GDP (0.996). The data set is obtained from the OECD Data. The average GFCF value (in 2009 million Dollars) was about $356 billion in 2018. The US had the highest gross fixed capital formation ($4.30 trillion), about 32% of the total of the OECD countries. Japan is the runner-up country that exceeds the $1 trillion threshold. Iceland has the lowest gross fixed capital formation value at $4.41 billion.

Labor force:

As discussed above, a strong correlation exists between capital stock and real GDP (0.966). The OECD Data provides the corresponding data set. In 2018, the US had the largest labor force, where 162.08 million people were employed, with an average unemployment rate of about 4 percent. The OECD countries' average employment is about 17.42 million, and Iceland and Luxembourg have the lowest labor force, less than 300,000.

Energy Consumption:

Energy consumption is also an important input variable in the manufacturing process and should be considered in the model to evaluate energy efficiency. As seen in Table 2, there is a strong correlation between primary energy consumption and the output variables. Primarily, there is a perfect correlation between fossil fuel consumption and CO_2 emission. The US Energy Information Administration (EIA) provides the corresponding data set. According to the EIA, the United States' primary energy consumption hit a historical high of 101.24 quadrillion BTU (British thermal units) in 2018. According to the EIA, 80 percent of consumed energy is from burning fossil fuels, the primary source of CO_2 emissions. The average energy consumption is about 6.48 quadrillion BTU. Estonia has the lowest energy consumption, 0.095 quadrillion BTU.

Output Variables

This study considers a desirable and undesirable output variable in the data set: (1) real GDP and (2) CO_2 Emissions, respectively. Table 3 presents the related data for these output variables.

Real GDP:

The real gross domestic product (GDP) is the most critical output variable to measure the states' eco-efficiency. The data set is collected from the OECD Data. As seen in Table 2, there is a strong correlation between the GDP and all the input variables. The average GDP of OECD countries is \$1.61 trillion, ranging from \$20.17 billion to \$20.53 trillion, which Iceland and the US generated, respectively. There are only 13 countries that exceed the \$1 trillion threshold.

CO_2 Emissions:

Table 2 shows a strong correlation between CO_2 emissions and all the other input variables, especially with primary energy consumption. Hence, this output contributes significantly to precisely assessing states' eco-efficiency and ecologically sustainable manufacturing performance. The data set is derived from the EIA's US States Energy Portal. According to the EIA, the states' average CO_2 Emission is about 307.87 million metric tons, and it ranges from 1.70 million and 4.91 billion tons, which Iceland and the US produced, respectively.

Table 3: Descriptive statistics of input and output variables in 2018

	Mean	StDev	Min	Median	Max
Gross fixed capital formation (GFCF)	355,965.42	723,130.42	4,405.60	119,175.10	4,295,134.00
Labor Force (LF)	17,427.20	29,070.73	203.90	5,143.95	162,075.00
Fossil Fuel Consumption (FFC)	5.19	13.52	0.10	1.05	82.80
Nuclear Power Consumption (NPC)	0.51	1.49	0.00	0.00	8.40
Renewable Energy Consumption (REC)	0.82	1.91	0.00	0.30	11.40
Gross Domestic Product (GDP)	1,610,204.74	3,409,904.68	20,174.90	490,727.65	20,533,058.00
CO_2 Emission (CE)	307.87	803.28	1.70	61.45	4,910.00

RESULTS AND DISCUSSIONS

DEA Results

This study quantitatively evaluates the eco-efficiency of the OECD countries using five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO2 emissions). The input-oriented CCR and BCC models are implemented to evaluate the states' technical efficiency and pure technical efficiency. In addition, the SBM model is developed to incorporate undesirable output variables directly—Figure 1 and Table 4 present the CCR, BCC, SBM, and scale efficiency score.

The input-oriented CCR model indicates eight countries: Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, reaching the maximum eco-efficiency score of 1.000 in 2018. The average eco-efficiency score is 0.863, which ranges between 0.545 and 1.000. In addition, according to the input-oriented CCR model, 24 countries' eco-efficiency score exceeds 0.80, and 17 exceed 0.90. On the other hand, Canada, Australia, and South Korea have the lowest eco-efficiency scores, less than 0.70, during the research focus period.

The input-oriented BCC model indicates that 24 countries reached the maximum eco-efficiency score of 1.000 in 2018. The average eco-efficiency score is 0.957, which ranges between 0.756 and 1.000. In addition, according to the input-oriented BCC model, 34 countries' eco-efficiency score exceeds 0.80, and 31 exceed 0.90. Similarly, during the research focus period, Chile, Hungary, New Zealand, and the Slovak Republic had the lowest eco-efficiency scores, less than 0.80.

The ratio of the technical efficiency of the CCR model and the pure technical efficiency of the BCC model calculates the scale efficiency scores for the output-oriented models. The mean scale efficiency score is 0.903. Eight countries, Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, are scale efficient, and 28 countries' scale efficiency scores exceed 80%, and 24 exceed 0.90.

Moreover, input-oriented SBM efficiency scores are computed to incorporate undesirable output variable. The results indicate that eight countries, Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, reached the maximum eco-efficiency score of 1.000 in 2018. The average eco-efficiency score is 0.711, which ranges between 0.318 and 1.000. In addition, according to the input-oriented SBM model, 12 countries' eco-efficiency score exceeds 0.80, and 9 exceed 0.90. On the other hand, Canada and South Korea have the lowest eco-efficiency scores, less than 0.40, during the research focus period.

The dual problem of the input-oriented CCR model provides states' current level of returns to scale information. For a given DMU, returns to scale are identified as increasing returns to scale (IRS) if the sum of the dual weights is less than 1, and it is identified as decreasing returns to scale (DRS) if the sum of the dual weights is greater than 1. Only eight eco-efficient countries, Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, are designated as CRS, indicating that they operated at the most excellent level of productive efficiency in 2018. On the other hand, the IRS is assigned to only two countries, Finland and Lithuania, implying that they might considerably increase their production by expanding their input variables. Twenty-eight countries are classed as DRS, suggesting that their production level improves less than the proportion growth in their inputs. Thirty countries benchmark Luxembourg due to the optimum input and output variables combination.

Figure 2 presents the average eco-efficiency scores of the OECD countries. There are eight eco-efficient countries: Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland. The average eco-efficiency score is 0.858, which ranges between 0.602 and 1.000. Twenty-four countries' eco-efficiency score exceeds 0.80, and 17 exceed 0.90.

Table 4: The Eco-efficiency Scores of the OECD countries

Country	CCR Efficiency	BCC Efficiency	Scale Efficiency	SBM Efficiency	RTS	Benchmark Country
Australia (AUS)	0.6531 (37)	0.9687 (27)	0.6742 (37)	0.5311 (29)	DRS	CRI, GRC, LUX.
Austria (AUT)	0.8762 (20)	0.9139 (31)	0.9587 (22)	0.7569 (18)	DRS	DNK, IRL, LUX.
Belgium (BEL)	0.7488 (29)	0.9396 (29)	0.7969 (30)	0.4653 (32)	DRS	CHE, CRI, LUX.
Canada (CAN)	0.6554 (36)	0.8403 (34)	0.7800 (32)	0.3376 (37)	DRS	GRC, LUX.
Chile (CHL)	0.7475 (30)	0.7902 (36)	0.9460 (23)	0.6331 (25)	DRS	CRI, LUX, LVA.
Colombia (COL)	0.8996 (18)	1.0000 (1)	0.8996 (25)	0.8245 (11)	DRS	CRI, IRL, LUX, LVA.
Costa Rica (CRI)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	CRI.
Czech Republic (CZE)	0.7385 (32)	0.9986 (25)	0.7395 (33)	0.5645 (27)	DRS	LUX, LVA.
Denmark (DNK)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	DNK.
Estonia (EST)	0.9360 (15)	1.0000 (1)	0.9360 (24)	0.8272 (10)	IRS	LUX, LVA.
Finland (FIN)	0.8333 (23)	0.8427 (33)	0.9888 (12)	0.5237 (30)	DRS	CRI, DNK, LUX.
France (FRA)	0.9689 (12)	1.0000 (1)	0.9689 (19)	0.7958 (13)	DRS	CHE, CRI, LUX, LVA.
Germany (DEU)	0.8690 (21)	1.0000 (1)	0.8690 (26)	0.5568 (28)	DRS	CRI, DNK, LUX.
Greece (GRC)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	GRC.
Hungary (HUN)	0.7691 (27)	0.7994 (35)	0.9621 (20)	0.4941 (31)	DRS	CHE, CRI, LUX, LVA.
Iceland (ISL)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	ISL.
Ireland (IRL)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	IRL.
Israel (ISR)	0.7869 (26)	1.0000 (1)	0.7869 (31)	0.7908 (17)	DRS	LUX, LVA.
Italy (ITA)	0.9827 (10)	1.0000 (1)	0.9827 (13)	0.9115 (9)	DRS	CRI, DNK, LUX.
Japan (JPN)	0.7005 (35)	1.0000 (1)	0.7005 (36)	0.4119 (36)	DRS	CHE, CRI, LUX, LVA.
Korea (KOR)	0.5452 (38)	1.0000 (1)	0.5452 (38)	0.3180 (38)	DRS	CRI, LUX, LVA.
Latvia (LVA)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	LVA.
Lithuania (LTU)	0.8827 (19)	0.8863 (32)	0.9959 (10)	0.7938 (15)	IRS	CRI, DNK, ISL, LUX.
Luxembourg (LUX)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	LUX.
Mexico (MEX)	0.8081 (24)	1.0000 (1)	0.8081 (28)	0.4534 (33)	DRS	CRI, LUX, LVA.
Netherlands (NLD)	0.7989 (25)	1.0000 (1)	0.7989 (29)	0.6447 (23)	DRS	CRI, GRC, LUX.
New Zealand (NZL)	0.7410 (31)	0.7557 (38)	0.9805 (15)	0.6362 (24)	DRS	CRI, DNK, LUX.
Norway (NOR)	0.9671 (13)	0.9933 (26)	0.9736 (18)	0.8045 (12)	DRS	DNK, IRL, ISL, LUX.
Poland (POL)	0.8649 (22)	1.0000 (1)	0.8649 (27)	0.7448 (20)	DRS	LUX.
Portugal (PRT)	0.9286 (16)	0.9506 (28)	0.9769 (17)	0.7957 (14)	DRS	CRI, GRC, LUX.
Slovak Republic (SVK)	0.7641 (28)	0.7718 (37)	0.9900 (11)	0.4476 (34)	DRS	CRI, GRC, LUX.
Slovenia (SVN)	0.9588 (14)	1.0000 (1)	0.9588 (21)	0.7163 (21)	DRS	LUX, LVA.
Spain (ESP)	0.9015 (17)	0.9211 (30)	0.9787 (16)	0.6233 (26)	DRS	CRI, DNK, LUX.
Sweden (SWE)	0.9986 (9)	1.0000 (1)	0.9986 (9)	0.7489 (19)	DRS	CHE, CRI, LUX.
Switzerland (CHE)	1.0000 (1)	1.0000 (1)	1.0000 (1)	1.0000 (1)	CRS	CHE.
Turkey (TUR)	0.7325 (34)	1.0000 (1)	0.7325 (35)	0.6518 (22)	DRS	DNK, IRL, LUX, LVA.
United Kingdom (GBR)	0.9823 (11)	1.0000 (1)	0.9823 (14)	0.7922 (16)	DRS	CHE, CRI, LUX, LVA.
United States (USA)	0.7378 (33)	1.0000 (1)	0.7378 (34)	0.4228 (35)	DRS	GRC, LUX.
Mean	0.8626	0.9572	0.9030	0.7110		
Standard Deviation	0.1250	0.0754	0.1193	0.2100		
Minimum	0.5452	0.7557	0.5452	0.3180		
First Quartile	0.7485	0.9350	0.7984	0.5293		
Second Quartile	0.8795	1.0000	0.9655	0.7469		
Third Quartile	0.9867	1.0000	0.9966	0.8483		
Maximum	1.0000	1.0000	1.0000	1.0000		

Figure 1: The Eco-efficiency Scores of the OECD countries

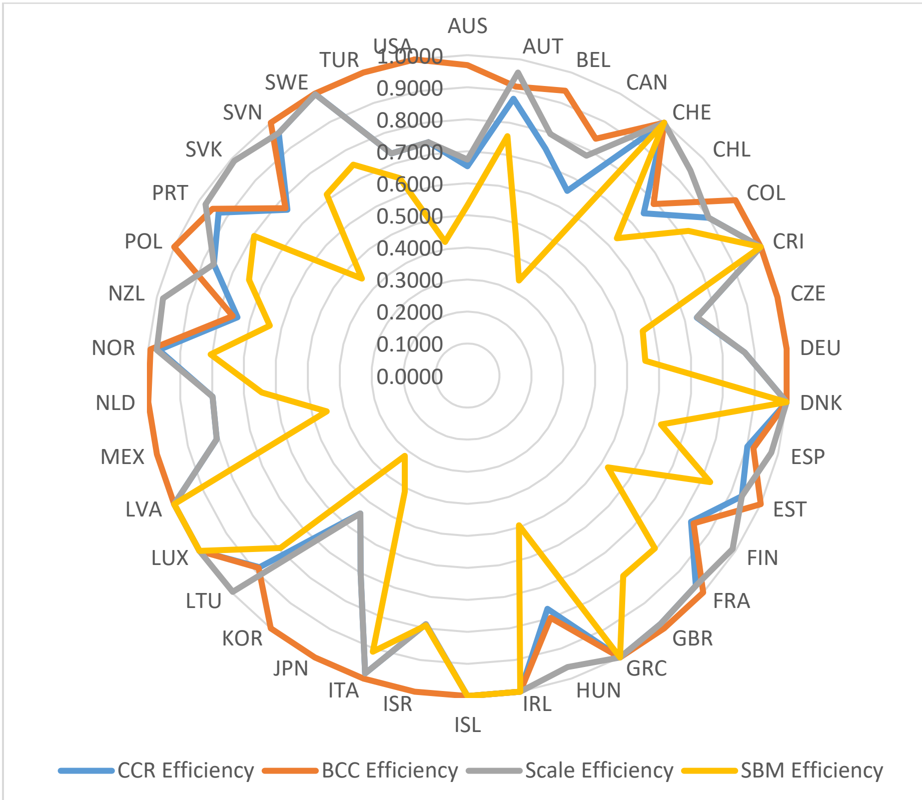


Figure 2: The Average Eco-efficiency Scores of the OECD countries

SUMMARY AND CONCLUSION

This study quantitatively evaluates the eco-efficiency of the OECD countries using five inputs (capital stock, labor force, fossil fuel consumption, nuclear power consumption, and renewable energy consumption), one desirable output (real GDP), and one undesirable output variable (CO2 emissions). The input-oriented CCR and BCC models are implemented to evaluate the states' technical efficiency and pure technical efficiency. In addition, the SBM model is developed to incorporate undesirable output variables directly.

Only eight eco-efficient countries, Costa Rica, Denmark, Greece, Iceland, Ireland, Latvia, Luxembourg, and Switzerland, are designated as CRS, indicating that they operated at the most excellent level of productive efficiency in 2018. On the other hand, the IRS is assigned to only two countries, Finland and Lithuania, implying that they might considerably increase their production by expanding their input variables. Twenty-eight countries are classed as DRS, suggesting that their production level improves less than the proportion growth in their inputs. Thirty countries benchmark Luxembourg due to the optimum input and output variables combination.

In conclusion, it is hoped that the findings of this study will enable policymakers to understand the latest relative eco-efficiency assessments of the OECD countries.

REFERENCES

Apergis, N., Aye, G.C., Barros, C.P., Gupta, R. and Wanke, P., 2015. Energy efficiency of selected OECD countries: A slacks based model with undesirable outputs. *Energy Economics*, *51*, pp.45-53.

Arrhenius, S. and Holden, E.S., 1897. On the Influence of Carbonic Acid in the Air upon the Temperature of the Earth. *Publications of the Astronomical Society of the Pacific*, *9*(54), pp.14-24.

Banker, R.D., Charnes, A. and Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*(9), pp.1078-1092.

Caves, D.W., Christensen, L.R. and Diewert, W.E., 1982. The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*, pp.1393-1414.

Charnes, A., Cooper, W.W. and Rhodes, E., 1978. Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, *2*(6), pp.429-444.

Choi, Y., Zhang, N. and Zhou, P., 2012. Efficiency and abatement costs of energy-related CO2 emissions in China: A slacks-based efficiency measure. *Applied Energy*, *98*, pp.198-208.

Demiral, E.E. and Sağlam, Ü., 2021. Eco-efficiency and Eco-productivity assessments of the states in the United States: A two-stage Non-parametric analysis. *Applied Energy*, *303*, p.117649.

Demiral, E.E. and Sağlam, Ü., 2023. Sustainable production assessment of the 50 US states. *Journal of Cleaner Production*, *419*, p.138086.

Fourier, M. 1827. Les temperatures du globe terrestre et des espaces planetaires. *Memoires de l'Academie Royale des Sciences de l'Institut de France*, *7*.

Iftikhar, Y., Wang, Z., Zhang, B. and Wang, B., 2018. Energy and CO2 emissions efficiency of major economies: A network DEA approach. *Energy*, *147*, pp.197-207.

IPCC, 2018: Summary for Policymakers. In: *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty* [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. *World Meteorological Organization, Geneva, Switzerland, 32 pp.*

Moutinho, V., Madaleno, M. and Robaina, M., 2017. The economic and environmental efficiency assessment in EU cross-country: Evidence from DEA and quantile regression approach. *Ecological Indicators*, *78*, pp.85-97.

Moutinho, V., Fuinhas, J.A., Marques, A.C. and Santiago, R., 2018. Assessing eco-efficiency through the DEA analysis and decoupling index in the Latin America countries. *Journal of Cleaner Production*, *205*, pp.512-524.

OECD (2023), Investment (GFCF) (indicator). doi: 10.1787/b6793677-en (Accessed on May 1, 2023)

OECD (2023), Air and GHG emissions (indicator). Doi: 10.1787/93d10cf7-en (Accessed on May 1, 2023)

Park, Y.S., Lim, S.H., Egilmez, G. and Szmerekovsky, J., 2018. Environmental efficiency assessment of US transport sector: A slack-based data envelopment analysis approach. *Transportation Research Part D: Transport and Environment*, *61*, pp.152-164.

Sağlam, Ü., 2017a. A two-stage data envelopment analysis model for efficiency assessments of 39 state's wind power in the United States. *Energy Conversion and Management*, 146, 52-67.

Sağlam, Ü., 2017b. Assessment of the productive efficiency of large wind farms in the United States: an application of two-stage data envelopment analysis. *Energy Conversion and Management,* 153, 188-214.

Sağlam, Ü., 2017c, November. The productive efficiency assessment of wind power generation in the United States. In *2017 Annual Meeting of the Decision Sciences Institute Proceedings, Washington DC*.

Sağlam, Ü., 2018a. A two-stage performance assessment of utility-scale wind farms in Texas using data envelopment analysis and Tobit models. *Journal of Cleaner Production*, *201*, pp.580-598.

Sağlam, Ü., 2018b, November. The Efficiency Assessment of Renewable Energy Sources with Data Envelopment Analysis. In *2018 Annual Meeting of the Decision Sciences Institute Proceedings, Chicago IL*.

Sağlam, Ü., 2019. The effects of electricity prices on productive efficiency of states' wind power performances in the United States. *Economics Bulletin*, AccessEcon, vol. 39(2), pages 866-875.

Sağlam, Ü., 2022, November. Eco-efficiency Assessment of the US States: A DEA-Based Malmquist Productivity Index Analysis. In *2022 Annual Meeting of the Decision Sciences Institute Proceedings, Houston TX*.

Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis*. European Journal of Operational Research*, 130(3), pp.498-509.

US Bureau of Labor Statistics – BLS – Economy at a Glance. Retrieved January 15, 2022, from https://www.bls.gov/eag/

US Department of Commerce – Bureau of Economic Analysis – Reginal Data, GDP and Personal Income. Retrieved January 15, 2022, from https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1&acrdn=1

US Energy Information Administration – EIA – US States Energy Portal. Retrieved January 15, 2022, from https://www.eia.gov/beta/states/overview

US Energy Information Administration – EIA – International Emissions. Retrieved January 15, 2022, from https://www.eia.gov/international/data/world/other-statistics/emissions

Wang, K., Yu, S. and Zhang, W., 2013. China's regional energy and environmental efficiency: A DEA window analysis based dynamic evaluation. *Mathematical and Computer Modelling*, *58*(5-6), pp.1117-1127.

Wang, C.N., Hsu, H.P., Wang, Y.H. and Nguyen, T.T., 2020. Eco-efficiency assessment for some European countries using slacks-based measure data envelopment analysis. *Applied Sciences*, *10*(5), p.1760.

Yeh, T.L., Chen, T.Y. and Lai, P.Y., 2010. A comparative study of energy utilization efficiency between Taiwan and China. *Energy Policy*, *38*(5), pp.2386-2394.

Zhou, P. and Ang, B.W., 2008. Linear programming models for measuring economy-wide energy efficiency performance. *Energy Policy*, *36*(8), pp.2911-2916.

Zhou, P., Ang, B.W. and Han, J.Y., 2010. Total factor carbon emission performance: a Malmquist index analysis. *Energy Economics*, *32*(1), pp.194-201.

DECISION SCIENCES INSTITUTE

EDI Utilization in Pharmacy Revenue Cycle and the Prioritization of Exploitation over Exploration

Derek Dubois
Ph.D. student, University of Rhode Island
Derek.Dubois@URI.edu

Dr. Mehmet Yalcin
Associate Professor, University of Rhode Island
MGYalcin@URI.edu

Dr. Muhammad Hasan Ashraf
Assistant Professor, California State University Long Beach
Hasan.Ashraf@CSULB.edu

ABSTRACT

Healthcare revenue cycle requires governance for industry standards maintained by meta-organizations that create coalitions between competing firms. While Electronic Data Interchange (EDI) enables significant operational efficiencies (exploitation), is there enough focus on the development of future standards (exploration)? Paradox Theory contextualizes persistent tensions between competing goals within firms. The present study utilizes empirical methods to explore practitioners' attitudes in identifying and categorizing the tensions. The goal is to build upon established exploration vs. exploitation literature, assessing how practitioners within the healthcare revenue cycle space can more adequately strategize for future needs effectively.

KEYWORDS: Pharmacy, Healthcare Economics, Supply Chain Management, Electronic Data Interchange (EDI), Paradox Theory

DECISION SCIENCES INSTITUTE

Effects of Gamification Training on Employees' Information Security Compliance

Venkata Chaitanya Kumar Suram
University of the Cumberlands
vsuram3418@ucumberlands.edu
Mary Lind
Louisiana State University Shreveport
Mary.lind@lsus.edu

ABSTRACT

The purpose of this study was to determine whether gamification training significantly influences employees' protection motivations and security compliance behaviors. Scholars have begun to suggest that gamification, or the integration of gaming elements into training programs, might positively affect employees' information security compliance. This study added to the body of knowledge on gamification, protection motivation theory, and information security policy compliance by investigating five research questions. The research questions independently examined the extent to which gamification influenced the relationships between the predictor variables of response efficacy, resource availability, self-efficacy, security policy attitude, and detection certainty and the outcome variable information security policy compliance intentions.

KEYWORDS: Information Security Compliance, Gamification, Protection Motivation Theory

INTRODUCTION

Increasing numbers of organizations use training on information security risks for employees and its effects on employees' security compliance behaviors (Karagiannis et al., 2020; Sharif & Ameen, 2021). Researchers and practitioners are starting to recognize the great potential of gamification in information technology and security (Sharif & Ameen, 2021). The behavioral aspects of cybersecurity have gained more attention in recent years because users can hinder the successful implementation of security measures based on their technical skill levels and attitudes toward information security (Park et al., 2019; Van Steen & Deeleman, 2021). Security awareness campaigns that merely provide security information need more engagement in the training by employees (Scholl, 2018), so cybersecurity professionals are designing games to facilitate cybersecurity training (Scholl, 2018; Silic & Lowry, 2020; Van Steen & Deeleman, 2021). This research explores the effectiveness of using gamified training on employees' Information Security Compliance Behaviors.

Security policy violations are the key concern of most organizations. Employees who fail to comply with security policies represent significant security risks (Theis et al., 2019; Wang et al., 2018). Employees' behavioral conduct contributes dramatically to security policy violations (Gratian et al., 2018; Mamonov & Benbunan-Fich, 2018). The most common security policies violated in contemporary business are information security policies (Cram et al., 2019).
All organizations need to secure the information. When organizations fail to manage their information security, organizational integrity could be compromised, and financial losses could occur (Alhayani et al., in press). Organizations have traditionally considered information security a technological problem with a technical solution, but this approach does not account for the risks associated with users' behaviors (Aurigemma & Mattson, 2018; Homoliak et al., 2019). Organizations seeking to manage behavioral risks must identify and measure insider threats to information assets and develop ways to manage those threats effectively (Alohali et al., 2018; Evans et al., 2019; Urueña López et al., 2019).

Information in various forms is the most critical asset for most organizations. Companies in the private and public sectors must systematically secure information, ensuring it remains confidential, accurate, and accessible (Gratian et al., 2018; Safa et al., 2018). Ensuring information security concerns the entire organization, not only management (Andersson et al., 2022). Organizations use information security policies to protect critical information assets (Chen et al., 2018). An information security policy explains what an employee should and should not do and outlines employees' responsibilities to secure information (Bauer et al., 2017; Chen et al., 2018). Pérez-González et al. (2019) stated that an information security policy gives management direction and support for protecting data assets. Paananen et al. (2020) noted that establishing standards is a good starting point for creating information security policies and enhancing information security in organizations.

Some scholars have begun to suggest that gamification, or the integration of gaming elements into training programs, might positively affect information security outcomes (Jawad & Tout, 2021; Ophoff & Dietz, 2019; Rintanalert & Luangsodsai, 2021). Rintanalert and Luangsodsai (2021) conducted an experimental study to examine the use of gamification among university students and found that gamification training did elicit more active participation. However, Rintanalert and Luangsodsai did not find significant improvement in information security awareness. Other studies that link gamification to cybersecurity and technology also focus on student populations, leaving a gap in the literature (Jawad & Tout, 2021; Ophoff & Dietz, 2019; Pan et al., 2017). Scholars have noted a need for more focused gamification research (Schöbel et al., 2021). It is presently unknown whether gamified training can improve security policy compliance among employees. The lack of knowledge regarding the ability of gamification to significantly influence employees' security policy compliance constitutes a research problem. The present study examined whether Indian employees' security policy compliance intentions differ significantly based on their training type (traditional vs. gamified).

The purpose of this study was to measure the relationship between the study's independent and dependent variables to determine whether gamification training significantly influences employees' protection motivations and security compliance behaviors. Response efficacy, resource availability, self-efficacy, security policy attitude, and detection certainty functioned as the study's independent variables. Employees' information security policy compliance intentions functioned as the study's dependent variable. The study addressed added to the body of knowledge on gamification, protection motivation theory, and information security policy compliance (Diesch et al., 2020; Koohang et al., 2020; Rintanalert & Luangsodsai, 2021; Schöbel et al., 2021). Gamification is becoming a popular approach in information security training because it increases employee engagement, but the effectiveness of the approach remains unclear (Cram et al., 2019; Rintanalert & Luangsodsai, 2021). Measuring whether employees who receive gamification training exhibit higher levels of information security policy compliance compared to employees who receive traditional training would enable organizations to tailor their training approach to secure their data effectively. More effective training programs would be beneficial, as many employees accidentally expose their employers to risk due to poor or ineffective security training (Homoliak et al., 2019; Prabhu & Thompson, 2020).
Research on information security compliance develops expertise in the research topic and increases career growth in this subject area (Moody et al., 2018). The research is also important because it provides appropriate security training mechanisms to employees for data security compliance (Aviles, 2017).

LITERATURE REVIEW

Protection motivation served as the study's theoretical framework. Protection motivation theory was first developed by Rogers (1975). Rogers noted that the desire to protect against threats was closely linked to fear appeals. Rogers suggested that when faced with a threat, individuals examine the threat from two perspectives: (a) how much of a threat a situation poses and (b) how well a person can mitigate the threat. In later research Rogers (1983) argued that individuals are motivated by a threat's perceived severity, perceived vulnerability to the threat, and the effectiveness of existing threat prevention methods.

Since protection motivation theory was first proposed in 1975, scholars have used the theory to investigate or explain behavior in many fields (Herath & Rao, 2009). Scholars have applied the theory in different ways, with some scholars using the research to examine personal risks and others using the theory to study organizational threats (Herath & Rao, 2009). Many scholars have used protection motivation to explore factors affecting information security policy compliance (Chang et al., 2018; Hassandoust & Techatassanasoontorn, 2020; Hina et al., 2019; Van Bavel et al., 2019; Verkijika, 2018).

The main variables associated with protection motivation have evolved. When Rogers (1975) first proposed protection motivation theory, it contained constructs labeled appraised severity, the expectation of exposure, and belief in the efficacy of coping response. As the theory evolved, the theoretical constructs and interrelationships changed. Herath and Rao's (2009) model was used as a basis for the present study. As a result, the main independent variables examined in the present study were response efficacy, resource availability, self-efficacy, security policy attitude, and detection certainty. Each of those variables was assumed to influence compliance intentions directly. Additionally, the significance of those relationships was assumed to differ significantly based on whether an employee received traditional information security policy training or gamified training.

Information Security

Many cyberattacks occur daily in organizations (Silic & Lowry, 2020) and these attacks can damage organizations' reputations and cost millions of dollars in financial damages. Most information security vulnerabilities, cyberattacks, risks, and viruses are due to a lack of security awareness among employees, workers, customers, and users (Silic & Lowry, 2020). Information technology security compliance offers techniques and procedures that encourage personnel to act more securely while working with organizational structures and information. Usually, information security compliance has three targets: (a) to mitigate or avoid security incidents and risks often caused by negligent employees, (b) to thwart criminal security conduct and computer abuse, and (c) to protect against employees' security behaviors (Silic & Lowry, 2020). This research study focuses on the significance of gamification training for information security compliance.

Employees and contractors are the primary sources of information breaches, and most security experts say insider threats are on the rise (Aurigemma & Mattson, 2018; Gratian et al., 2018; Homoliak et al., 2019; Mamonov & Benbunan-Fich, 2018). Privileged users have access to sensitive information posing the most significant risk followed by consultants and contractors and regular employees (Spears & Barki, 2017). Some of the reasons for the breeches are:

- Insufficient data protection strategies and solutions
- Wide variety of devices with access to sensitive information
- Proliferation of sensitive information moving past the firewall on mobile devices
- More accessibility to the network by regular employees, contractors, and partners
- Complexity of technology
- Growing use of cloud apps.

Many employee-related breaches are accidental. Negligent employees can ignore organizational decisions to update security patches (AlKalbani et al., 2017). The Equifax breach, which uncovered the sensitive data of tens of millions of Americans, resulted from a single employee's mistake. It is hard to believe that careless employees are not malicious; however, the Equifax case proves that maliciousness and neglect are not always the same (AlKalbani et al., 2017).

Employee Information Security Awareness

Uninformed employees are one of the greatest threats to information security. Education and attention are essential to mitigate the danger; however, organizations can find it challenging to enforce their vital front-line training and awareness applications. Gamification offers an approach to mitigate this situation, ensuring that knowledge of safety information is added to the employee training program at the right time. The 21st century has included technological improvements that have helped organizations flourish and work faster and more efficiently. For the company to be competitive, it must adapt to evolving security threats.

With the digital enterprise, there is a higher threat of cybercrime. In line with Cram et al. (2019), there was a 38% growth in public organization data security breaches in 2016 compared to 2014. Organizations must plan and operate secure digital systems to exchange data and funds where companies have adopted information security compliance methods to control data usage. Investing in information security improves organizational security mechanisms that help safeguard information (Siponen et al., 2020). Compliance measures governing data generation and protection then facilitate agreements between the organization and its stakeholders regarding information security (Cram et al., 2019).

Avgar et al. (2018) noted that data security must preserve data integrity, confidentiality, and accessibility. However, information cannot be preserved without compliance with the standards that guide its preservation. During the last decade, there was the perception that data security was primarily a technical concern. However, more recently, scholars and practitioners have begun to view data security's human (i.e., social and personal) aspects (Aurigemma & Mattson, 2018; Njenga, 2017; Siponen & Vance, 2010). Changing perspectives have resulted in new strategies, guidelines, and consciousness applications that support security compliance. In step with Siponen et al. (2020), the failure to prevent security breaches suggests that a corporation is not complying with its security policies. Research has ascertained that almost half of security breaches emanate from inside the enterprise (Aurigemma & Mattson, 2018; Homoliak et al., 2019; Siponen & Vance, 2010). The sources of these data breaches suggest that organizations need to focus more on stopping security breaches via compliance.

Siponen et al. (2020) noted that training to introduce employees to security policies is critical to ensuring data security. The training programs should identify procedures and sanctions for negligent or non-compliant behaviors. Other factors can affect organizational security and compliance behaviors, including employees' job roles and support from top management (Siponen et al., 2020). For these reasons, managers should also receive the necessary training so that compliant security behaviors can be modeled from the top down within an organization.

Several theories explain individuals' information security compliance behaviors. Spears and Barki (2017) explained that institution theory focuses on the pressures an organization puts on individuals to comply with organizational policies. Organizations attain legitimacy by conforming to the external expectancies of their stakeholders (Spears & Barki, 2017). The legitimacy that an organization seeks may be won via making strategic responses to the pressures from outside entities (Spears & Barki, 2017). However, organizations must remain aware of external factors and how they influence their structure and operation (Siponen & Vance, 2010).
External pressures force businesses to monitor information security compliance. These pressures include normative, mimetic, and coercive pressures (Spears & Barki, 2017). Coercive pressures like regulations cause organizations to undertake information security measures to avoid penalties and fines associated with security breaches. Coercive pressures require organizations to adhere to laws and policies established by governing authorities (Siponen et al., 2020). Normative pressures arise within an organization (Siponen et al., 2020). Normative pressures include expectations of management that organizations will keep data secure. Mimetic pressures are the third type of pressure that can affect how organizations approach information security compliance. Mimetic pressures originate from observing peers and

competitors (Siponen & Vance, 2010). Organizations often institute security policies like other companies in their field to bolster their legitimacy.

When examining the use of gamification in response to information security and the various pressures that influence organizational security behaviors, the most relevant type of pressure is mimetic pressure. More organizations are using gamification elements in employee training and awareness programs. Siponen and Vance (2010) suggested that employees' information security compliance increases when employees become more security conscious, and awareness is a critical factor in robust information security compliance. Governments can create guidelines and rules that pressure businesses to prioritize data security, but compliance may not be automatic unless employees pay attention and value the role of information security. Thus, gamification can be a strategic tool for raising employee awareness.

Employees' Information Security Compliance

Work on information security compliance is impossible without assigning responsibility for compliance (Johnson et al., 2016). It is critical to understand the role of the board, executives, managers, and employees working on sensitive information. Employee ownership in an organization at the grass-root level improves employees' information security attitudes and compliance (Johnson et al., 2016).

Adequate information safety benefits corporations and employees. Employees want to comply with well-designed data protection practices, so organizations should focus on promoting awareness and understanding of data handling procedures and information security policies. Corporations can effectively guard against hackers and protect clients, contacts, and suppliers by monitoring security needs and regularly updating their information handling protocols to address emerging threats (Yang et al., 2017). When all employees are alert to information security risks, it reflects the strength of the organization's commitment to security.

Failing to adopt proper information security protocols can cause problems for organizations and even lead to data security breaches. All organizations that store and process information digitally are at risk from cyberattacks (Cram et al., 2019). Management groups must be proactive in maintaining defense against these threats and communicating security-related behavioral expectations to their employees. Organizations sometimes forget temporary employees or contractors when addressing data security. However, organizations must ensure that every employee of the business is trained in information security awareness (Wang et al., 2019).

Enabling an environment where employees feel a sense of shared responsibility and accountability for data security protection is important (Hussain et al., 2018). Compliance culture critically affects employees because it means an organization takes cyber-protection seriously (Hussain et al., 2018). It is vital that the push for safety and vigilance starts from the top and works down, empowering all team members (Hussain et al., 2018). Encouraging a compliance culture and providing employees with regular training demonstrates that organizations will hold individuals responsible for personal data security efforts, improving compliance. Investing in a compliance culture can improve cyber protection by ensuring that the skills learned in security training become part of everyday operational practices (Yang et al., 2017).

Knowledge sharing, collaboration, intervention, and experience significantly impact employees' behavior in compliance with organizational information security rules (Cram et al., 2019). However, research indicates that attachment does not significantly affect employees' behavior toward information security policy compliance (Cram et al., 2019). Research has also shown that commitment and private norms affect personnel beliefs (Cram et al., 2019). Mindset towards compliance with information security organizational guidelines also dramatically impacts information security compliance (Spears & Barki, 2017). Organizations seeking to address human factors related to information security must account for these considerations.

Consequences of Noncompliance

Cyberattacks and poor information security compliance cost organizations millions in daily losses (Moody et al., 2018). These losses have created intense interest in information security policies and employee compliance among scholars and practitioners (AlGhamdi et al., 2020; Alotaibi et al., 2019; Brazevich et al., 2020; Diesch et al., 2020; Koohang et al., 2020). Most organizations are developing information security policies to address risks, but few studies have been conducted on how training affects employee awareness or compliance intentions where scholars like Homoliak et al. (2019) and Prabhu and Thompson (2020) have acknowledged that inadequate security training contributes to poor security behaviours. However, most studies focused on training covered strategies to counteract social engineering (Schaab et al., 2017), addressed the role of internal audits (Stafford et al., 2018), and the identification of security awareness factors (Diesch et al., 2020).

Understanding the reasons for security policy violations is critical because these violations can cost organizations millions of dollars (Moody et al., 2018). Information is one of the most vital assets in many organizations, and almost all modern organizations have established security programs to protect against data theft and attack. However, ensuring data security is more challenging than ever before. Cheap labour, job insecurity, new and poorly trained workers can influence employee loyalty and consequent employee behaviours leading to poor security policy compliance (Wang et al., 2019)

Gamification

Most organizations already use reward structures to promote worker engagement and policy compliance, but gamification represents a new approach to training that can improve motivation through interest and hedonic benefits (Rodwald, 2019; Safa et al., 2018). Gamification allows companies to cut through the noise of day-to-day activities, subtly improve behavioral outcomes, enhance personnel engagement and motivation, and improve the company's information security (Sailer et al., 2017). Gamification presents employees with demanding situations that inspire them to become emotionally invested in the quality of their work. Gamified solutions leverage game design, loyalty program design, and behavioral change to clear up non-public and business data security issues (Hussain et al., 2018; Tsohou et al., 2020). Gamification is particularly effective as research indicates that making rewards variable increases dopamine and supercharges interest through the anticipation of delight (Hussain et al., 2018).

Offering extrinsic credit can powerfully reinforce a behavioral change and encourage employee engagement (Sailer et al., 2017). One of the most significant errors training developers make when using gamification is to concentrate on a single ultimate reward and include insufficient numbers of small rewards (Sailer et al., 2017). The rewards that most effectively motivate users tend to be the smaller rewards. Continuous smaller rewards help prevent employees from losing interest before they finish training (Hussain et al., 2018). Participants often rationalize their refusal to follow security policy guidelines, and poor compliance can negate password protection measures (Siponen et al., 2020). Linking these behaviors to small challenges and rewards can improve compliance.

Organizations must understand employee goals and organizational needs related to data security for information security training to be successful. Additionally, organizations should understand how gamification techniques work before implementing them to ensure they are intrinsically tied to an organization's core data security mission (Siponen et al., 2020). Personnel should feel fulfilled by seeing their progress measured visibly. However, if there is a disconnect with gamified training, employees can mistrust the process, rendering the training unsuccessful (Hussain et al., 2018)

Practical Training Using Gamification

The integration of game factors into an organization's administrative center requires management to understand the challenges and benefits of game inclusion and how game design reflects physical gameplay and related work actions (Cram et al., 2019). Through gamification, new employees are encouraged to participate and improve their knowledge of behavioural patterns associated with information security policy compliance. Video games encourage modeled behaviour and can show employees how to approach demanding situations (Cram et al., 2019).

Gamification makes the online guides more understandable for newcomers, presenting them with an efficient learning platform as they progress and gain more knowledge. Video games assist in triggering discussions on unique study concepts or critical training points (Hussain et al., 2018). Games' adaptable and robust nature means they can be customized and tailored to organizational goals or learner skill levels. Novices should feel encouraged and feel they can meet organizational milestones on the way to the targeted behaviour (Chen, 2019).
Skill level is a critical consideration in the use of gamification. Inexperienced individuals might feel intimidated by gamified training courses. Participants could feel conscious about their scores and grades because of their social reputation, leading to extended participation in the e-studying program (Wang et al., 2019).

Unfortunately, even if a learner is encouraged to comply with a policy due to gamified training, actual compliance behavior will not arise if that individual does not have the necessary skills. Motivation resulting from gamification encourages inexperienced individuals to work harder to complete tasks in set timeframes.

While moving between training topics, new employees need to know what to expect. Using gamification to help inexperienced employees feel motivated to gain knowledge rather than chastising them for their inexperience is an essential benefit of a nontraditional approach to training (Cram et al., 2019). Implementing game-based knowledge as part of work training helps increase learners' interest, enterprise productivity, and step-by-step shift from the classroom-based learning environment to self-directed mastering and collaborative activities (Choi, 2016).

Continuous Training

The benefits of worker compliance training and improvement are widely known. As the organization grows and the person develops with it, knowledge needs to be exchanged, and employee skillsets must be regularly updated. Organizations must develop complete employee education programs, and improvement plans must be developed continuously, revised, and tailored for new conditions, markets, and demanding situations if companies wish to remain competitive (Sailer et al., 2017). Information security regulations must also be modified with marketplace adjustments and technological development. More modern and complicated rules are evolving daily, affecting all aspects of the organization, including ethics and the employees' training programs (Cram et al., 2019).

Gamification in information security training uses elements drawn from games to make obligations more attractive for employees and customers and improve the chances of meeting organizational goals (AlKalbani et al., 2017). Gamification allows monotonous tasks to be more enjoyable and incorporating gaming elements in training programs has many benefits, including financial savings, performance improvements, and increased employee engagement (AlKalbani et al., 2017). Organizations use game mechanics like reward points, contests, and social recognition to stimulate and encourage employees to follow information security policies and procedures (Silic & Lowry, 2020).

Organizations continue to invest in gamified designs, with some estimates suggesting marketplace growth of gamification to be 48% in 2019 (Silic & Lowry, 2020). Silic and Lowry (2020) noted that failed gamification is rare, and organizations hoping to avoid failure should tailor their gamification design to their organizational goals. Design elements and approaches must be carefully selected to increase gamified system success. Badges and leaderboards are not effective for every training objective, but carefully laying out gamified designs is an excellent way to encourage desired behavior (Silic & Lowry, 2020).

Gamification designs offer healthy ways to strengthen information security policies, reassure worried customers, and engage employees. Organizations provide a behavioral framework through gamification training and highlighting principles for improving data handling. Organizations hope that gamification will produce more nuanced knowledge of the organizations' information systems (Liu et al., 2017). The dialogue of gamification design concepts and pattern research questions has established that varied disciplinary theories ar essential to research. Instead of viewing gamified systems as a separate research technology, organizations should consider gamification an information system design paradigm. Several existing literature streams associated with gamification include intrinsically motivating, hedonic, and persuasive systems, technology adoption and use, media characteristics, online and virtual groups, and user-generated content (Baxter et al., 2016).

Many gamification rewards, including badges, are intangible and considered irrelevant. However, intangible rewards differ from tangible extrinsic rewards often studied in economics and the intrinsic rewards addressed by psychological theories. Despite the differences, rewards are unmistakably a crucial part of digital video games and a precursor to intrinsic motivation. Hence, one of the theoretical challenges is conceptualizing those rewards and perhaps enlarging the inherent extrinsic framework of gamification (AlKalbani et al., 2017).

Gamified information systems replicate a growing convergence between instrumental and experiential structures. Using technologies and ideas from virtual gaming, gamification proponents posit that actual-global structures should be reengineered to be more attractive and productive (Liu et al., 2017). Researchers face challenges and blessings when developing or integrating gamification training into existing systems. Gamification offers progressive ideas for system layouts for tracking progress, novel perspectives for research, and a new pedagogy to entice employees security compliance (Silic & Lowry, 2020).

Benefits of Gamification

Gamification is the technique wherein the expertise and experience received from gaming theory and flow theory have been used in a non-gaming context. Gamification became applied for the first time during the cold war to improve productiveness (Huotari & Hamari, 2017). In 1984, gamification was used in a commercial enterprise context to encourage employees via frequent feedback, gaming features, and personal goals (Kamel et al., 2017). Gamification allows organizations to improve their employees' engagement levels by utilizing numerous game design features (Liu et al., 2017). Some researchers have recommended that goals, storytelling, rewards, and appreciation are the principal aspects of gamification (Ruiz-Alba et al., 2019). These aspects encourage curiosity, interest, and engagement. They also address user challenges and enable contributors to provide training sessions and workshops (Ruiz-Alba et al., 2019)

Gamification is becoming one of the most preferred training methodologies, allowing organizations to increase employees' innovation, productivity, knowledge, talents, reviews, and learning procedures (Sailer et al., 2017). Gamified training primarily uses innovative thoughts and gaming techniques in a non-entertainment way to enhance education and work capabilities. Gamification benefits organizations because it enables employees to bolster their productiveness, improve their organizational engagement and involvement, develop problem-solving skills and innovative attitudes, and strengthen communication methods (Yang et al., 2017).

Gamification bolsters employee engagement by introducing several innovative dynamics (Huotari & Hamari, 2017). Organizations that use gamification techniques in training sessions find that employees exhibit significant improvement in demonstrating the required abilities

(Huotari & Hamari, 2017). Gamification also increases employee interest and involvement in training classes and workshops (Kamel et al., 2017).

Gamification is considered a beneficial method of presenting a positive corporate image (Liu et al., 2017). Organizations that want to use gamification more effectively must keep the process simple, enticing, and unique, as these traits improve employee engagement levels. Gamification's success relies mainly on employee involvement, gaming techniques and methods, and motivation (Ruiz-Alba et al., 2019). The rewards offered are not considered just rewards; they provide a way to uplift employees and improve their skills.

When organizations use gamification, they work to make existing tasks more innovative and fun, like video games. The development in information technology has increased cybercrimes and terrorism which can have a strong negative impact on an organization's reputation and the information and data saved in the organization's servers, affecting personnel, customers, and the organization itself (Sailer et al., 2017). The accelerated numbers of breaches, threats, risks, and vulnerabilities call for IT corporations to become more progressive, productive, and dependable (Straub & Welke, 2018; Yang et al., 2017). For this reason, organizations must offer training sessions and workshops to improve their employee's talents and skillset. To perceive and address the numerous attacks, threats, risks, and vulnerabilities, employees must kno about information security to guard their privacy and information against intruders (Huotari & Hamari, 2017).

Employees are the most vulnerable access point for breaches and assaults, so they must be well-trained (Gratian et al., 2018; Mamonov & Benbunan-Fich, 2018). Hence, employee training is considered distinctly influential in any organization because they deal directly with the organization's data and information security infrastructure. When employees are provided with data knowledge, practice, and experience, they can make appropriate decisions to prevent breaches that put the organization at risk (Spears & Barki, 2017).

Gamification uses interactive principles to make cybersecurity attractive, enjoyable, and informative (Avgar et al., 2018). Employees learn techniques to prevent data leakages and follow the rules through enjoyable and engaging training sessions. Following policies and preventing data breaches are foundational to security compliance. Using targeted training methods, organizations can improve information security compliance and instill a pro-security culture within the entire corporation.

Using Gamification to Promote Information Security Compliance

Flow theory is considered beneficial for explaining the technique wherein gamification usage can enhance individuals' abilities and skills. The flow experience in games creates a balance between employees' talents and cyber security situations (Marinho et al., 2019). While a worker can feel exhausted and bored with education sessions, gamification can revive interest in the training session as they move to a higher level in the training (Ruiz-Alba et al., 2019). Sailer et al. (2017) noted that the flow principle addresses how a person engages in an activity that helps improve their cognitive abilities. The engagement includes the unique feeling of managing, being completely focused on the operation, playing the scenario, and having the essential harmony among the abilities and the mission completed (Huotari & Hamari, 2017). Consequently, organizations can foster intrinsic motivation through gamification by making security policy compliance more enjoyable (Sailer et al., 2017).

Since employees are the most critical points of vulnerability for data breaches, cyberattacks, and assaults (Gratian et al., 2018; Mamonov & Benbunan-Fich, 2018), gamification training uses challenges and rewards to make cybersecurity attractive, engaging, and informative, so employees learn to follow information security rules and prevent breaches (Ruiz-Alba et al., 2019). Using appropriate cyber security training methods enhances information security compliance and supports heightened organizational security. In this way, gamification increases the connection between information security compliance and data security (Siponen et al., 2020).

Current research indicates that personalizing security training content can make learning more relevant and understandable, and mixing this with sensible exercises improves security behaviors. E-learning is one of the easy and effective integrations of gamification. Game techniques and methods can be implemented in the training process as activities to achieve specific organizational goals, increase learners' motivation to complete them, and engage employees in a friendly competitive environment with other employees.

RESEARCH DESIGN and FINDINGS

This study used a comparative, inferential, quantitative research approach to determine whether gamification training could be used to improve information security policy compliance. The study's specific purpose was to determine whether gamification training significantly influences employees' protection motivations and security compliance behaviors. The population of interest in this study consisted of employees working for large companies in India that use information security policies to protect organizational data and IT infrastructure. India has a corporate culture where workers adhere to rules and policies (Deshpande & Farley, 1999) so this context is an interesting test to see if gamification makes a difference.

After identifying the target population, a sampling frame was established. The selection process focused on an organization that embraced data systems security and was willing to consider gamification training to bolster information security policy compliance. India is a diverse country with many regions, so it was important to choose a company carefully. The company needed to have employee's representative of the national population and management willing to participate in the study. Also, the firm had just begun to implement gamified cyber security training and was willing to administer the gamified training to half of its employees while providing the traditional cyber training to the other employees. The gamification cyber training involved cybersecurity challenges with rewards. Using proper sampling methods in quantitative research is critical to ensure that a study's results are generalizable to the target population (Fisher et al., 2018). Participants were informed about the study via a notice distributed by their employer. The notice was distributed to all the employees who received either traditional or gamified information security policy training. Employees could choose to visit the link, and there was no undue influence. Participants learned that their personal information would not be collected or shared. Demographic data were not collected from participants to ensure individuals could not be identified based on individual survey responses.

The protection motivation instrument validated by Herath and Rao (2009) served as this study's data collection instrument. Herath and Rao developed their instrument based on several theories, including protection motivation theory and deterrence theory. The purpose of Herath and Rao's study was to help explain how behavioral, organizational, and environmental factors influenced information security policy compliance. Herath and Rao (2009) examined a range of independent variables. The survey consisted of 43 items. Each item was measured on a 7-point Likert scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*).

The analysis consisted of descriptive statistics and multivariate analysis of variance (MANOVA) analysis to determine the differences between two groups of employees. The sample consisted of two groups of employees from an organization in India. One group of employees participated in gamified information security compliance training, while the other group received traditional training. The analysis aimed to determine whether gamification training influences the relationships between protection motivation variables and security compliance behaviors.

The study was guided by five hypotheses:

H_01: There is no statistically significant predictive relationship between response efficacy and information security policy compliance intentions based on gamification training participation.
H_A1: There is a statistically significant predictive relationship between response efficacy and information security policy compliance intentions based on gamification training participation.
H_02: There is no statistically significant predictive relationship between resource availability and information security policy compliance intentions based on gamification training participation.
H_A2: There is a statistically significant predictive relationship between resource availability and information security policy compliance intentions based on gamification training participation.
H_03: There is no statistically significant predictive relationship between self-efficacy and information security policy compliance intentions based on gamification training participation.
H_A3: There is a statistically significant predictive relationship between self-efficacy and information security policy compliance intentions based on gamification training participation
H_04: There is no statistically significant predictive relationship between security policy attitude and information security policy compliance intentions based on gamification training participation.
H_A4: There is a statistically significant predictive relationship between security policy attitude and information security policy compliance intentions based on gamification training participation.
H_05: There is no statistically significant predictive relationship between detection certainty and information security policy compliance intentions based on gamification training participation.
H_A5: There is a statistically significant predictive relationship between detection certainty and information security policy compliance intentions based on gamification training participation.

Participants and Research Setting

The sample consisted of N = 414 employees from an organization in India. Participants were divided into two groups. One group received gamification training meant to improve information security compliance, and the other group received traditional training. Demographic data on age and gender were not collected, meaning the group characteristics could not be differentiated based on these common factors. However, the employee pool was homogenous, and no major differences were expected between the two groups related to age, gender, or work experience.

The research setting was a company in India with over 500 employees. The company embraced technology, and the company's employees were expected to comply with established security policies. The company employed traditional security policy compliance training methods, but a subset of employees was also trained using new gamification methods. The purpose of the new training was to improve compliance outcomes among employees. The organization granted permission to survey their employees with the understanding that they would benefit from the results when it was determined whether gamification training could improve information security policy compliance. The inclusion criteria required all participants to work full-time at the company.

Analyses of Research Questions

This research investigated the extent to which the use of gamification training influences employees' protection motivations and security compliance behaviors? This question was answered by testing five hypothesis pairs. All the study's hypotheses were tested using

MANOVA analysis. The data were coded to assign a value of 0 to traditional training and 1 to gamification training. All participants were asked what kind of training they received. The sample was then divided into two groups based on participants' responses. The final totals of the two groups were *n* = 228 for traditional training and *n* = 186 for gamification training.

Descriptive statistics were calculated after separating the data into two groups. The descriptive statistics included mean (*M*) and standard deviation (*SD*) values for each survey item. Table 1 presents the mean and standard deviation values for each survey item among participants that received traditional training. Table 2 presents the values for participants that received gamification training. Participants receiving gamification training reported higher mean scores for all survey items. The lowest mean score for participants with traditional training was *M* = 1.65 for Response efficacy 2. The lowest mean score for participants with gamification training was *M* = 2.96 for Response efficacy 1. High scores were similarly different with traditional training resulting in *M* = 2.54 for Resource availability 1 and gamification training resulting in *M* = 4.85 for Compliance intention 1.

Table 1

Descriptive Statistics for Employees with Traditional Training

Item	*M*	*SD*	Range
Response efficacy 1	2.40	0.820	1-5
Response efficacy 2	1.65	0.496	1-5
Response efficacy 3	2.02	0.883	1-5
Resource availability 1	2.54	1.084	1-5
Resource availability 2	2.46	0.603	1-5
Resource availability 3	1.85	0.647	1-5
Resource availability 4	2.22	1.094	1-5
Resource availability 5	2.45	0.781	1-5
Self-efficacy 1	2.00	0.660	1-5
Self-efficacy 2	1.84	0.377	1-5
Self-efficacy 3	2.02	0.883	1-5
Security policy attitude 1	2.21	0.890	1-5
Security policy attitude 2	2.52	0.826	1-5
Security policy attitude 3	1.89	0.661	1-5
Detection certainty 1	2.04	0.753	1-5

Detection certainty 2	2.02	0.640	1-5
Compliance intention 1	2.06	0.724	1-5
Compliance intention 2	2.41	0.501	1-5
Compliance intention 3	2.51	0.736	1-5

Note.N = 228.

Table 2

Descriptive Statistics for Employees with Gamification Training

Item	*M*	*SD*	Range
Response efficacy 1	2.96	0.944	1-5
Response efficacy 2	4.00	0.000	1-5
Response efficacy 3	4.14	0.993	1-5
Resource availability 1	4.08	0.292	1-5
Resource availability 2	3.44	0.497	1-5
Resource availability 3	4.44	0.549	1-5
Resource availability 4	4.42	0.566	1-5
Resource availability 5	3.96	0.949	1-5
Self-efficacy 1	4.01	0.073	1-5
Self-efficacy 2	4.02	1.000	1-5
Self-efficacy 3	4.81	0.574	1-5
Security policy attitude 1	3.63	0.637	1-5
Security policy attitude 2	4.39	0.633	1-5
Security policy attitude 3	3.63	0.663	1-5
Detection certainty 1	3.10	0.314	1-5
Detection certainty 2	4.40	0.643	1-5
Compliance intention 1	4.85	0.486	1-5
Compliance intention 2	4.34	0.624	1-5
Compliance intention 3	3.48	0.552	1-5

Note. N = 186.

Reliability was tested using the total dataset (*N* = 414). Detection certainty had a low-reliability score of 0.62, but that is likely because Cronbach's alpha calculations are susceptible to low scores when there are fewer items. Detection certainty might have had a higher reliability score if the subscale contained more than two items. Based on Cronbach's alpha scores, the data were considered reliable for analysis. Mahalanobis distance was used to test the linear relationships between the five independent variables of response efficacy, resource availability, self-efficacy, security policy attitude, and detection certainty and the dependent variable of compliance intention (McLachlan, 1999). The results of the test for multivariate normality. Tabachnik & Fidell (2007) identified the maximum critical value for Mahalanobis distance as 20.52. The maximum Mahalanobis distance value in the present study was 17.242, meaning that the assumption of normality was supported. Box's test of equality of covariance matrices was used to test the homogeneity for the dependent variables. The results of Box's *M* test confirmed that the dependent variables' observed covariance matrices were not homogenous between groups. While this violated the assumption of equal variance, the benefit of using a MANOVA analysis is that four analysis methods can be used to test the hypotheses (Wilk's lambda, Hotelling's trace, Pillai's trace, and Roy's largest root). When the homogeneity assumption is violated, Pillai's trace can still provide a robust evaluation of hypotheses using MANOVA (Ateş et al., 2019; Olson, 1974). Pillai's trace was used as a multivariate test to determine if the independent variables significantly affected the dependent variable of security policy compliance intentions. Table 3 presents the results of Pillai's trace for each variable. As seen in the Table 3, each independent variable significantly influenced participants' compliance intentions. The use of a MANOVA analysis and Pillai's trace was necessary because the study sought to determine whether gamification vs. traditional training affected the relationships between the independent variables and the dependent variable. After examining Pillai's trace, the next step was then to test for univariate influences.

The multivariate testing using Pillai's trace confirmed the significant effects of the independent variables of response efficacy, resource availability, self-efficacy, security policy attitude, and detection certainty on the dependent variable of compliance intention. Thus, it was important to use univariate testing to investigate the effect differences between the two groups (traditional vs. gamified training). Gamification acted as an additional variable affecting the relationships between the independent and dependent variables. This was the reason that a MANOVA analysis was used rather than an ANOVA analysis. To understand the effects of gamification, a between-subject effects test is necessary.

Table 3

Pillai's Trace

Effect	Value	*F*	Hypothesis *df*	Error *df*	Sig.
Intercept	0.684	218.545	4.000	404.000	<0.001
Response efficacy	0.807	423.044	4.000	404.000	<0.001
Resource availability	0.633	173.941	4.000	404.000	<0.001
Self-efficacy	0.659	194.899	4.000	404.000	<0.001
Security policy attitude	0.540	118.657	4.000	404.000	<0.001
Detection certainty	0.846	553.533	4.000	404.000	<0.001
Gamification	0.712	249.440	4.000	404.000	<0.001

Table 4 presents the results of the between-subject effects testing. In Table 4, the *F*-statistic and the *p*-value show whether significant differences exist between the gamification and traditional training groups. The univariate testing indicated that all the independent variables have statistically significant predictive relationships with the information security compliance intentions. Additionally, the testing confirmed that gamification training significantly improved compliance intentions compared to traditional training. Table 4 answers the research questions by accounting for several things. First, the multivariate approach ensures that the analysis does not overestimate the effect of one variable by not accounting for other variables in the model. That concern was why a single variable-based ANOVA was inappropriate for answering the study's research questions (Akbay et al., 2019). Second, a between-group test was necessary to understand the difference between the variable effects on the two groups (i.e., gamified vs. non-gamified). Here the *F*-statistic and the significance (*p*-value) reported in Table 4 answer the research questions by indicating whether significant differences exist between the groups. The null hypotheses can be rejected because the significance values are all below 0.001.

Table 4

Test of Between-Subject Effects

Source	Type III *SS*	*df*	M^2	*F*	Sig.	Partial eta^2
Corrected model	407.718[a]	6	67.953	1516.957	<0.001	0.957
Intercept	5.999	1	5.999	133.913	<0.001	0.248
Response efficacy	1.941	1	1.941	43.329	<0.001	0.096
Resource availability	3.131	1	3.131	69.900	<0.001	0.147
Self-efficacy	11.812	1	11.812	263.695	<0.001	0.393
Security policy attitude	3.729	1	3.729	83.242	<0.001	0.170
Detection certainty	21.115	1	21.115	471.355	<0.001	0.537
Gamification	7.521	1	7.521	167.907	<0.001	0.292
Error	18.232	407	0.045			
Total	4612.846	414				
Corrected total	425.950	413				

Note. R^2 = 0.957 (Adjusted R^2 = 0.957)

Findings

Table 5 presents a summary of the hypothesis testing results. As indicated in Table 5, all the independent variables were statistically significant predictors of compliance intentions when gamification training was compared to traditional security compliance training. The findings mean that the links between protection motivations were significantly stronger among participants who engaged in gamification training compared to participants who received traditional training.

Table 5

Hypothesis Summary

Hypothesis	Variable relationship	Result
1	Response efficacy > compliance intentions	Reject the Null
2	Resource availability > compliance intentions	Reject the Null
3	Self-efficacy > compliance intentions	Reject the Null
4	Security policy attitude > compliance intentions	Reject the Null
5	Detection certainty > compliance intentions	Reject the Null

SUMMARY and IMPLICATIONS

This study aimed to determine whether gamification training significantly influences employees' protection motivations and security compliance behaviors. Information security policies are critically important to data integrity, accessibility, and confidentiality (Avgar et al., 2018). Many scholars have identified employees as the primary risk to data security for not adhering to security policies (Theis et al., 2019; Wang et al., 2018). Organizations typically adopt technical solutions to combat cybersecurity but fail to address behavioural risks (Aurigemma & Mattson, 2018; Homoliak et al., 2019). One way to address behavioral risks is by developing information security policies and implementing training programs to raise employees' security awareness (Bauer et al., 2017; Chen et al., 2018). The study's specific purpose was to determine whether gamification training significantly influences employees' protection motivations and security compliance behaviors.

Practical Assessment of Research Questions

The results of the data analysis indicated that in each case, the relationships between the independent predictor variables and the dependent outcome variable (compliance intentions) were significant. Additionally, the between-groups effect test demonstrated that participants who used gamification training reported higher scores in relation to information security compliance intentions. As a result of the analysis, all the null hypotheses were rejected.

The first research question assessed the relationship between response efficacy and compliance intentions. Response efficacy refers to employee perceptions that their organization's information security policies effectively protect data confidentiality, availability, and integrity (Herath & Rao, 2009). The findings for the first research question aligned with protection motivation theory and supported previous research. Several studies highlighted the importance of response-efficacy as a predictor of security policy compliance intentions (Chang et al., 2018; Hassandoust & Techatassanasoontorn, 2020; Koohang et al., 2021; Koohang et al., 2020; Liu et al., 2017; Rajab & Eydgahi, 2019; van Bavel et al., 2019). Additionally, research has indicated that gamification links to organizational resources, response efficacy, and information security policy behavior (Alqahtani & Kavakli-Thorne, 2020; Huotari & Hamari, 2017; Silic & Lowry, 2020). Comparing the present study's findings with previous research demonstrated alignment and supported future research using protection motivation theory.

The second research question examined the relationship between resource availability and compliance intentions. Resource availability refers to whether an organization provides employees with information and support for information security policies (Herath & Rao, 2009). The findings for the second research question aligned with protection motivation theory and supported previous research. Several studies highlighted the importance of resource availability as a predictor of security policy compliance intentions (Balozian & Leidner, 2017; Jalali & Kaiser, 2018; Kumar & Goyal, 2019; Shuaib et al., 2019). Additionally, research has indicated that gamification links to resource availability and allocation and information security policy behavior (Dincelli & Chengalur-Smith, 2020; Mota et al., 2021; Sharif & Ameen, 2021). Comparing the present study's findings with previous research demonstrated alignment and supported future research using protection motivation theory.

The third research question addressed the relationship between self-efficacy and compliance intentions. Self-efficacy refers to an employee's ability to protect data by complying with information security policies (Herath & Rao, 2009). Self-efficacy is associated with skills and skill development, so there were parallels in the literature related to gamification that supported the promotion of self-efficacy. Several studies highlighted the importance of self-efficacy as a predictor of security policy compliance intentions (Hassandoust & Techatassanasoontorn, 2020; Hina et al., 2019; Liu et al., 2017; Mamonov & Benbunan-Fich, 2018; Van Bavel et al., 2019; Verkijika, 2018). Additionally, research has indicated that gamification links to improved skills, self-efficacy, and information security policy behavior (Huotari & Hamari, 2017; Marinho et al.,2019; Yang et al., 2017). Comparing the present study's findings with previous research demonstrated alignment and supported future research using protection motivation theory.

The fourth research question examined the relationship between security policy attitude and compliance intentions. Security policy attitude refers to employee perceptions regarding the importance, benefits, or helpfulness of security policies and practices (Herath & Rao, 2009). The findings for the fourth research question aligned with protection motivation theory and supported previous research. Several studies highlighted the importance of security policy attitude as a predictor of security policy compliance intentions (Ali et al., 2021; Amankwa et al., 2018; Koohang et al., 2020; Ormond et al., 2019; Paliszkiewicz, 2019). Additionally, research has indicated that gamification links to employee attitudes and the effects attitudes have on information security policy behaviours (Dincelli & Chengalur-Smith, 2020; Sharif & Ameen, 2020; Wu et al., 2021; Yang et al., 2017). Comparing the present study's findings with previous research demonstrated alignment and supported future research using protection motivation theory.

The fifth research question addressed the relationship between detection certainty and compliance intentions. Detection certainty refers to whether employees are monitored for security policy violations and if employees believe they will face the consequences for violating security policies (Herath & Rao, 2009). The findings for the fifth research question aligned with protection motivation theory and supported previous research. Several studies highlighted the importance of detection certainty as a predictor of security policy compliance intentions (Ali et al., 2021; Karjalainen et al., 2020; Merhi & Ahluwalia, 2019; Van Slyke & Belanger, 2020). Additionally, research has indicated that gamification links to detection certainty, fear appeals, and information security policy behavior (Creado & Ramteke, 2020; Ophoff & Dietz, 2019; Rodwald, 2019). Comparing the present study's findings with previous research demonstrated alignment and supported future research using protection motivation theory.

Implications for Future Study

This study's finding was that organizations should use gamification in training to increase employee engagement. The results from the analysis indicated that gamification training produced stronger relationships between all the protection motivation factors and employees' behavioral intentions to comply with information security policies. Finding ways to help employees engage in the information security process has benefits (Smith et al., 2016; Spanellis et al., 2020). Hussain et al. (2018) reported that gamification could improve employee motivation, engagement, and retention. This study's results supported previous research and showed that the benefits of gamification extend to improved information security policy compliance. The extensive literature identifying the benefits of gamification, supported by the present study's findings demonstrated that gamification is a valuable training and management tool.

REFERENCES

Akbay, L., Akbay, T., Erol, O., & Kilinç, M. (2019). Inadvertent use of ANOVA in educational research: ANOVA is not a surrogate for MANOVA. *Journal of Measurement and Evluation in Education and Psychology, 10*(3), 302–314. http://dx.doi.org/10.21031/epod.524511

Ali, R. F., Dominic, P. D. D., Ali, S. E. A., Rehman, M., & Sohail, A. (2021). Information security behavior and information security policy compliance: A systematic literature review for identifying the transformation process from noncompliance to compliance. *Applied Sciences, 11*(8), Article 3383. https://doi.org/10.3390/app11083383

AlGhamdi, S., Win, K. T., & Vlahu-Gjorgievska, E. (2020). Information security governance challenges and critical success factors: Systematic review. *Computers & Security, 99*(12), Article 102030. https://doi.org/10.1016/j.cose.2020.102030

Alhayani, B., Mohammed, H. J., Chaloob, I. Z., & Ahmed, J. S. (in press). Effectiveness of artificial intelligence techniques against cyber security risks appliestothe IT industry. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.02.531

AlKalbani, A., Deng, H., Kam, B., & Zhang, X. (2017). Information security compliance in organizations: An institutional perspective. *Data and Information Management, 1*(2), 104–114. https://doi.org/10.1515/dim-2017-0006

Alohali, M., Clarke, N., Li, F., & Furnell, S. (2018). Identifying and predicting the factors affecting end-users risk-taking behavior. *Information and Computer Security*, *26*(3), 306–326. https://doi.org/10.1108/ICS-03-2018-0037

Alotaibi, M. J., Furnell, S., & Clarke, N. (2019). A framework for reporting and dealing with end- user security policy compliance. *Information and Computer Security*, *27*(1), 2–25. https://doi.org/10.1108/ICS-12-2017-0097

Alqahtani, H., & Kavakli-Thorne, M. (2020). Does decision-making style predict individuals' cybersecurity avoidance behaviour? In A. Moallem (Eds.), *HCI for cybersecurity, privacy, and trust: Lecture notes in computer science* (pp. 32–50). Springer. https://doi.org/10.1007/978-3-030-50309-3_3

Amankwa, E., Loock, M., & Kritzinger, E. (2018). Establishing information security policy compliance culture in organizations. *Information and Computer Security, 26*(4), 420–436. https://doi.org/10.1108/ICS-09-2017-0063

Andersson, A., Hedström, K., & Karlsson, F. (2022). Standardizing information security: A structurational analysis. *Information & Management, 59*(3), Article 103623. https://doi.org/10.1016/j.im.2022.103623

Ateş, C., Kaymaz, Ö., Kale, H. E., & Tekindal, M. A. (2019). Comparison of test statistics of nonnormal and unbalanced samples for multivariate analysis of variance in terms of type-I error rates. *Computational and Mathematical Methods in Medicine, 2019*, Article 2173638. https://doi.org/10.1155/2019/2173638

Aurigemma, S., & Mattson, T. (2018). Exploring the effect of uncertainty avoidance on taking voluntary protective security actions. *Computers and Security*, *73*(3), 219–234. https://doi.org/10.1016/j.cose.2017.11.001

Avgar, A., Tambe, P., & Hitt, L. M. (2018). Built to learn: How work practices affect employee learning during healthcare information technology implementation. *MIS Quarterly, 42*(2), 645–659. http://dx.doi.org/10.25300/MISQ/2018/13668

Balozian, P., & Leidner, D. (2017). Review of IS security policy compliance: Toward the building blocks of an IS security theory. *The DATA BASE for Advances in Information Systems, 48*(3), 11–43. https://laur.lau.edu.lb:8443/xmlui/bitstream/handle/10725/6154/Review.pdf? sequence=1&isAllowed=y

Bauer, S., Bernroider, E. W. N., & Chudzikowski, K. (2017). Prevention is better than cure! Designing information security awareness programs to overcome users' non-compliance with information security policies in banks. *Computers & Security, 68*(7), 145–159. https://doi.org/10.1016/j.cose.2017.04.009

Baxter, R. J., Holderness, D. K., Jr., & Wood, D. A. (2016). Applying basic gamification techniques to IT compliance training: Evidence from the lab and field. *Journal of Information Systems, 30*(3), 119–133. http://dx.doi.org/10.2308/isys-51341

Brazevich, D. S., Safronova, Z. S., Kosheleva, T. N., & Biryukova, A. V. (2020). Analysis of the problems of ensuring information security in terms of the contemporary society. *Open Journal of Social Sciences, 8*(2), Article 98599. http://www.scirp.org/journal/Paperabs.aspx?PaperID=98599

Chang, S.-H., Hsu, H.-M., Li, Y., & Hsu, J. S.-C. (2018). The influence of information security stress on security policy compliance: A protection motivation theory perspective. *PACIS 2018 Proceedings, 2018*, Article 250. https://aisel.aisnet.org/pacis2018/250

Chen, X., Wu, D., Chen, L., & Teng, J. K. L. L. (2018). Sanction severity and employees' information security policy compliance: Investigating mediating, moderating, and control variables. *Information and Management*, *55*(8), 1049–1060. https://doi.org/10.1016/j.im.2018.05.011

Chen, Y. (2019). Exploring design guidelines of using user-centered design in gamification development: A Delphi study. *International Journal of Human-Computer Interaction, 35*(13), 1170–1181. https://doi.org/10.1080/10447318.2018.1514823

Choi, M. (2016). Leadership of information security manager on the effectiveness of information systems security for secure sustainable computing. *Information and Management, 8*(7), Article 638. https://doi.org/10.3390/su8070638

Cram, W. A., D'Arcy, J., & Proudfoot, J. G. (2019). Seeing the forest and the trees: A meta- analysis of the antecedents to information security policy compliance. *MIS Quarterly, 43*(2), 525–554. https://doi.org/10.25300/MISQ/2019/15117

Creado, Y., & Ramteke, V. (2020). Active cyber defence strategies and techniques for banks and financial institutions. *Journal of Financial Crime, 27*(3), 771–780. https://doi.org/10.1108/JFC-01-2020-0008

Deshpande, R., & Farley, J. U. (1999). Corporate culture and market orientation: Comparing Indian and Japanese firms. *Journal of International Marketing,* 7(4), 111-127. https://doi.org/10.1177/1069031X9900700407

Diesch, R., Pfaff, M., & Krcmar, H. (2020). A comprehensive model of information security factors for decision-makers. *Computers & Security, 92*(5), Article 101747. https://doi.org/10.1016/j.cose.2020.101747

Dincelli, E., & Chengalur-Smith, I. (2020). Choose your own training adventure: Designing a gamified SETA artifact for improving information security and privacy through interactive storytelling. *European Journal of Information Sytems, 29*(6), 669–687. https://doi.org/10.1080/0960085X.2020.1797546

Evans, M., He, Y., Maglaras, L., & Janicke, H. (2019). HEART-IS: A novel technique for evaluating human error-related information security incidents. *Computers and Security*, *80*(5), 74–89. https://doi.org/10.1016/j.cose.2018.09.002

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *PNAS, 115*(27), E6106–E6115. http://www.pnas.org/cgi/doi/10.1073/pnas.1711978115

Gratian, M., Bandi, S., Cukier, M., Dykstra, J., & Ginther, A. (2018). Correlating human traits and cyber security behavior intentions. *Computers and Security*, *73*(3), 345–358. https://doi.org/10.1016/j.cose.2017.11.015

Hassandoust, F., & Techatassanasoontorn, A. A. (2020). Understanding users' information ecurity awareness and intentions: A full nomology of protection motivation theory. In Benson & J. McAlhaney (Eds.), Cyber influence and cognitive threats (pp. 129–143). Academic Press. https://doi.org/10.1016/B978-0-12-819204-7.00007-5'

Herath, T., & Rao, H. R. (2009). Protection motivation and deterrence: A framework for security policy compliance in organizations. *European Journal of Information Systems, 18*, 106–125. http://dx.doi.org/10.1057/ejis.2009.6

Hina, S., Selvam, D. D. D. P., & Lowry, P. B. (2019). Institutional governance and protection motivation: Theoretical insights into shaping employees' security compliance behavior in higher education institutions in the developing world. *Computers & Security, 87*(11), Article 101594. https://doi.org/10.1016/j.cose.2019.101594

Homoliak, I., Toffalini, F., Guarnizo, J., Elovici, Y., & Ochoa, M. (2019). Insight into insiders and IT: A survey of insider threat taxonomies, analysis, modelling, and countermeasures. *ACM Computing Surveys, 52*(2), Article 8. https://doi.org/10.1145/3303771

Hu, S., Hsu, C., & Zhou, Z. (2021). Security education, training, and awareness programs: Literature review. *Journal of Computer Information Systems.* Advance online publication. https://doi.org/10.1080/08874417.2021.1913671

Huang, B., & Hew, K. F. (2021). Using gamification to design courses: Lessons learned in a three-year design-based study. *Journal of Educational Technology & Society, 24*(1), 44–63. https://www.jstor.org/stable/26977856

Huotari, K., & Hamari, J. (2017). A definition for gamification: Anchoring gamification in the service marketing literature. *Journal of Systems and Software, 27*(1), 21–31. https://doi.org/10.1007/s12525-015-0212-z

Hussain, S., Qazi, S., Ahmed, R. R., Streimikiene, D., & Vveinhardt, J. (2018). Employee management: Evidence from gamification techniques. *Montenegrin Journal of Economics, 14*(4),97–107. http://dx.doi.org/10.14254/1800-5845/2018.14-4.7

Jalali, M. S., & Kaiser, J. P. (2018). Cybersecurity in hospitals: A systematic organizational perspective. *Journal of Medical Internet Research, 20*(5), e10059. https://doi.org/10.2196/10059

Jawad, H. M., & Tout, S. (2021). Gamifying computer science education for Z generation. *Information, 12*(11), Article 453. https://doi.org/10.3390/info12110453

Kamel, M. M., Watfa, M. K., Lobo, B., & Sobh, D. (2017). Is enterprise gamification being cannibalized by its brand? *IEEE Transactions on Professional Communication, 60(2)*, 147–164. http://dx.doi.org/10.1109/TPC.2017.2656598

Karagiannis, S., Papaioannou, T., Magkos, E., & Tsohou, A. (2020). Game-based information security/privacy education and awareness: Theory and practice. In M. Themistocleous, M. Papadaki, & M. M. Kamal (Eds.), *Information systems: EMCIS 2020: Lecture notes in business information processing* (Vol. 402, pp. 509–525). Springer. https://doi.org/10.1007/978-3-030-63396-7_34

Karjalainen, M., Siponen, M., & Sarker, S. (2020). Toward a stage theory of the development of employees' information security behavior. *Computers & Security, 93*(6), Article 101782. https://doi.org/10.1016/j.cose.2020.101782

Koohang, A., Nord, J. H., Sandoval, Z. V., & Paliszkiewicz, J. (2021). Reliability, validity, and strength of a unified model for information security policy compliance. *Journal of Computer Information Systems, 61*(2), 99–107. https://doi.org/10.1080/08874417.2020.1779151

Koohang, A., Nowak, A., Paliszkiewicz, J., & Nord, J. H. (2020). Information security policy compliance: Leadership, trust, role values, and awareness. *Journal of Computer Information Systems, 60*(1), 1–8. https://doi.org/10.1080/08874417.2019.1668738

Kumar, R., & Goyal, R. (2019). Assurance of data security and privacy in the cloud: A three- dimensional perspective. *Software Quality Professional, 21*(2), 7–26. https://www.proquest.com/openview/66665ae705198eee3f12585401b42130/1?pq- origsite=gscholar&cbl=25782

Liu, D., Santhanam, R., & Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS Quarterly, 41(4)*, 1011–1034. http://dx.doi.org/10.25300/MISQ/2017/41.4.01

Mamonov, S., & Benbunan-Fich, R. (2018). The impact of information security threat awareness on privacy-protective behaviors. *Computers in Human Behavior*, *83*(6), 32–44. https://doi.org/10.1016/j.chb.2018.01.028

Marinho, A., Bittencourt, I. I., dos Santos, W. O., & Dermeval, D. (2019). Does gamification improve the flow experience in the classroom? An analysis of gamer types in collaborative and competitive settings. *Journal of Information Systems Education, 27*(2), 40–68.http://dx.doi.org/10.5753/rbie.2019.27.02.40

McLachlan, G. J. (1999). Mahalanobis distance. *Resonance, 1999*(6), 20–26. https://www.ias.ac.in/article/fulltext/reso/004/06/0020-0026

Merhi, M., & Ahluwalia, P. (2019). Examining the impact of deterrence factors and norms on resistance to information systems security. *Computers in Human Behavior, 92*(3), 37–46. https://doi.org/10.1016/j.chb.2018.10.031

Moody, G. D., Siponen, M., & Pahnila, S. (2018). Toward a unified model of information security policy compliance. *MIS Quarterly, 42*(1), 285–311. http://dx.doi.org/10.25300/MISQ/2018/13853

Mota, J. S., Tives, H. A., & Canedo, E. D. (2021). Tool for measuring productivity in software development teams. *Information, 12*(10), Article 396. https://doi.org/10.3390/info12100396

Njenga, K. (2017). Understanding internal information systems security policy violations as paradoxes. *Interdisciplinary Journal of Information, Knowledge & Management, 12*, 1–15. https://doi.org/10.28945/3639

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*(348), 894–907. https://doi.org/10.1080/01621459.1974.10480224

Ophoff, J., & Dietz, F. (2019). Using gamification to improve information security behavior: A password strength experiment. In L. Drevin & M. Theocharidou (Eds.), *Information security education: Education in proactive information security* (Vol. 557, pp. 157–169). Springer. https://doi.org/10.1007/978-3-030-23451-5_12

Ormond, D., Warkentin, M., & Crossler, R. E. (2019). Integrating cognition with an affective lens to better understand information security policy compliance. *Journal of the Association for Information Systems, 20*(12), Article 4. https://doi.org/10.17705/1jais.00586

Paananen, H., Lapke, M., & Siponen, M. (2020). State of the art in information security policy development. *Computers & Security, 88*(1), Article 101608. https://doi.org/10.1016/j.cose.2019.101608

Paliszkiewicz, J. (2019). Information security policy compliance: Leadership and trust. *Journal of Computer Information Systems, 59*(3), 211–217. https://doi.org/10.1080/08874417.2019.1571459

Pan, Y., Mishra, S., & Schwartz, D. (2017, June). Gamifying cybersecurity course content for entry-level students. Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, Ohio. https://peer.asee.org/27736

Park, M., Oh, H., & Lee, K. (2019). Security risk measurement for information leakage in IoT- based smart homes from a situational awareness perspective. *Sensors, 19*, Article s19092148. http://dx.doi.org/10.3390/s19092148

Pérez-González, D., Preciado, S. T., & Solana-Gonzalez, P. (2019). Organizational practices as antecedents of the information security management performance: An empirical investigation. *Information Technology & People, 32*(5), 1262–1275. https://doi.org/10.1108/ITP-06-2018-0261

Prabhu, S., & Thompson, N. (2020). *A unified classification model of insider threats to information security* [Paper presentation]. ACIS 2020: 31st Australasian Conference on Information Systems, Wellington, New Zealand. http://hdl.handle.net/20.500.11937/81763

Rajab, M., & Eydgahi, A. (2019). Evaluating the explanatory power of theoretical frameworks on intention to comply with information security policies in higher education. *Computers & Security, 80*(1), 211–223. https://doi.org/10.1016/j.cose.2018.09.016

Rintanalert, T., & Luangsodsai, A. (2021). Gamifying digital learning platform for information security awareness. In Y. M. Huang, C. F. Lai, & T. Rocha (Eds.), *Innovative technologies and learning: ICITL 2021: Lecture notes in computer science* (Vol. 13117, pp. 352–364). Springer. https://doi.org/10.1007/978-3-030-91540-7_37

Rodwald, P. (2019). Using gamification and fear appeal instead of password strength meters to increase password entropy. *Scientific Journal of Polish Naval Academy, 60*(2), 17–33. http://dx.doi.org/10.2478/sjpna-2019-0010

Rogers, R. W. (1983). Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection and motivation. In *Social psychophysiology: A sourcebook* (pp. 153-176). Guilford.

Ruiz-Alba, J. L., Soares, A., Rodríguez-Molina, M. A., & Banoun, A. (2019). Gamification and entrepreneurial intentions. *Journal of Small Business and Enterprise Development, 26*(5), 661–683.http://dx.doi.org/10.1108/JSBED-09-2018-0266

Safa, N. S., Maple, C., Watson, T., & Von Solms, R. (2018). Motivation and opportunity-based model to reduce information security insider threats in organizations. *Journal of Information Security and Applications*, *40*(6), 247–257. https://doi.org/10.1016/j.jisa.2017.11.001

Sailer, M., Hense, J., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of specific game design elements on psychological need satisfaction *Computers in Human Behavior, 69*(3),371–380. http://dx.doi.org/10.1016/j.chb.2016.12.033

Schaab, P., Beckers, K., & Pape, S. (2017). Social engineering defence mechanisms and counteracting training strategies. *Information and Computer Security*, *25*(2), 206–222. https://doi.org/10.1108/ICS-04-2017-0022

Schöbel, S., Schmidt-Kraepelin, M., Janson, A., & Sunyaev, A. (2021). Adaptive and personalized gamification designs: Call for action and future research. *AIS Transactions on Human-Computer Interaction, 13*(4), 479–494. https://doi.org/10.17705/1thci.00158

Scholl, M. (2018). Awareness of information security. *Systemics, Cybernetics, and Informatics, 16*(4), 80–89. http://iiisci.org/journal/PDV/sci/pdfs/IP059LL18.pdf

Sharif, K. H., & Ameen, S. Y. (2020). A review of security awareness approaches with special emphasis on gamification. In *2020 International Conference on Advanced Science and Engineering* (pp. 151–156). IEEE. https://doi.org/10.1109/ICOASE51841.2020.9436595

Sharif, K. H., & Ameen, S. Y. (2021, December 4–6). *A review on gamification for information security training*[Paper presentation].2021 International Conference of Modern Trends in Information and Communication Technology Industry (MTICTI),Sana'a, Yemen. https://doi.org/10.1109/MTICTI53925.2021.9664771

Shuaib, M., Samad, A., Alam, S., & Siddiqui, S. T. (2019). Why adopting the cloud is still a challenge? A review on issues and challenges for cloud migration in organizations. In Y. C. Hu, S. Tiwari, K. Mishra, & M. Trivedi, (Eds.), *Ambient communications and computer systems: Advances in intelligent systems and computing* (Vol. 904, pp. 387–399). Springer. https://doi.org/10.1007/978-981-13-5934-7_35

Silic, M., & Lowry, P. B. (2020). Using design-science-based gamification to improve organizational security training and compliance. *Journal of Management Information Systems, 37*(1), 129–161.http://dx.doi.org/10.1080/07421222.2019.1705512

Siponen, M., Puhakainen, P., & Vance, A. (2020). Can individuals' naturalization techniques be overcome? A field experiment on password policy. *Computers &Security, 88*(1), Article 101617. https://doi.org/10.1016/j.cose.2019.101617

Siponen, M., & Vance, A. (2010). Neutralization: New insights into the problem of employee information systems security policy violations. *MIS Quarterly, 34*(3), 487–502. http://dx.doi.org/10.2307/25750688

Smith, H. J., Milberg, S. J., & Burke, S. J. (2016). Information privacy Measuring individuals' concerns about organizational practices. *MIS Quarterly, 20*(2), 167–196.http://dx.doi.org/10.2307/249477

Spanellis, A., Dörfler, V., & MacBryde, J. (2020). Investigating the potential for using gamification to empower knowledge workers. *Expert Systems with Applications, 160*, Article 113694. https://doi.org/10.1016/j.eswa.2020.113694

Spears, J. L., & Barki, H. (2017). Employers' perspectives on new information technology technicians. *MIS Quarterly, 34*(3), 929–945. https://doi.org/10.1108/E-02-2017-0019

Spector, P. E. (2019). Do not cross me: Optimizing the use of cross-sectional designs. *Journal of Business and Psychology, 34*, 125–137. https://doi.org/10.1007/s10869-018-09613-8

Stafford, T., Deitz, G., & Li, Y. (2018). The role of internal audit and user training in information security policy compliance. *Managerial Auditing Journal*, *33*(4), 410–424. https://doi.org/10.1108/MAJ-07-2017-1596

Straub, D. W., & Welke, R. J. (2018). Coping with systems risk: Security planning models for management decision making. *MIS Quarterly, 22*(4), 441–469.http://dx.doi.org/10.2307/249551

Tabachnik, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Pearson.

Tsohou, A., Siponen, M., & Newman, M. (2020). How does information technology-based service degradation influence consumers' use of services? An information technology- based service degradation decision theory. *Journal of Information Technology, 35*(1), 2–24. https://doi.org/10.1177%2F0268396219856019

Urueña López, A., Mateo, F., Navío-Marco, J., Martínez-Martínez, J. M., Gómez-Sanchís, J., Vila-Francés, J., & José Serrano-López, A. (2019). Analysis of computer user behavior, security incidents and fraud using self-organizing maps. *Computers and Security*, *83*(6), 38–51. https://doi.org/10.1016/j.cose.2019.01.009

Van Bavel, R., Rodríguez-Priego, N., Vila, J., & Briggs, P. (2019). Using protection motivation theory in the design of nudges to improve online security behavior. *International Journal of Human-Computer Studies, 123*(3), 29–39. https://doi.org/10.1016/j.ijhcs.2018.11.003

Van Slyke, C., & Belanger, F. (2020). Explaining the interactions of humans and artifacts in insider security behaviors: The mangle of practice perspective. *Computers & Security, 99*(12), Article 102064. https://doi.org/10.1016/j.cose.2020.102064

Van Steen, T., & Deeleman, J. R. A. (2021). Successful gamification of cybersecurity training. *Cyberpsychology, Behavior, and Social Networking, 24*(9), 593–598. https://doi.org/10.1089/cyber.2020.0526

Verkijika, S. F. (2018). Understanding smartphone security behaviors: An extension of the protection motivation theory with anticipated regret. *Computers & Security, 77*(8), 860–870. https://doi.org/10.1016/j.cose.2018.03.008

Wang, J., Shan, Z., Gupta, M., & Rao, H. R. (2019). A longitudinal study of unauthorized access attempts on information systems: The role of opportunity contexts. *MIS Quarterly, 43(2)*, 601–622. http://dx.doi.org/10.25300/MISQ/2019/14751

Wang, X., Tan, Q., Shi, J., Su, S., & Wang, M. (2018). Insider threat detection using characterizing user behavior. *2018 IEEE Third International Conference on Data Science in Cyberspace, 2018*, 476–482. https://doi.org/10.1109/DSC.2018.00077

Wu, T., Tien, K., Hsu, W., & Wen, F. (2021). Assessing the effects of gamification on enhancing information security awareness knowledge. *Applied Sciences, 11*(10), Article 9266. https://doi.org/10.3390/app11199266

Yang, Y., Asaad, Y., & Dwivedi, Y. (2017). Examining the impact of gamification on the intention of engagement and brand attitude in the marketing context. *Computers in Human Behavior, 73*, 459–469

DECISION SCIENCES INSTITUTE

Emotion Detection in Texts: Efficiency Comparison of Different Classifiers

Debangi Choudhury, M.Sc.
Ravenshaw University, Odisha, India
debangi993@gmail.com

Alok Ranjan Tripathy, Ph.D.
Ravenshaw University, Odisha, India
aloktripathy.cs@ravenshawuniversity.ac.in

Meera Behera, Ph.D.
Georgian Court University, Lakewood, New Jersey
mbehera@georgian.edu

ABSTRACT

Emotion is a key component of communication which takes many forms, including facial expressions, gestures, voice, and writing. Text based emotion detection (TBED) is a relatively recent categorization with applications in marketing, social media, and eLearning. To extract texts based on different types of emotions, a text-based emotion detection method employs a machine learning (ML) approach. The aim of this paper is to test the efficiency of different machine learning classifiers used in emotion detection system. On our corpus, we used Naive Bayes classifier, Logistic regression, KNN, Decision tree, SVM, and Random Forest as baseline algorithms. Results of the tests and assessment reveals that SVM works with maximum efficiency.

KEYWORDS: Emotion detection, Machine learning classifiers, Natural language processing, and Affective computing

INTRODUCTION

Emotion is basically defined as any powerful feeling and includes experience, physiological behavior, cognitions, and conceptualization (Ortony et al.,1988; Garett B., 2009). Definitions of emotions and their categories have long been a subject of study. No one has been able to answer the question "How many different emotions do we have?" until today. Human emotions are complicated. Getting one feeling from another is a difficult task. Because it is dependent on a number of elements. It is difficult to determine one's emotional condition. It might be tough to tell where one feeling stops and another initiates. However, there is one exception. One can experience numerous emotions at the same time, making it even more difficult to discern (Ramalingam et al., 2018). Fear, happiness, furious, sadness, and surprise are all emotions that people experience. In the brain, each emotion has its own neural network and appropriate responses in terms of behavior. These fundamental emotions that compose/combine to generate complex emotions. For example, excitement can be a mix of joy and surprise, whilst disgust might be a mix of rage and misery. When it comes to recognizing emotion in text, there are several obstacles to overcome (Haji & Wu, 2010). It may seem difficult to infer someone's psychological response from a text document they have written, yet it is essential because the majority of textual information are not only direct and involve emotion in words, but also stem from the usage of sentimental words. The interpretation of concepts' meanings and how ideas interact as they are laid down in the document's language. The key to establish a link between people and machines is emotion recognition in text (Cowie et al., 2001). Voice, facial expression, and textual material can all be used to direct emotions, which are denoted as speech, facial, and text-based emotions, respectively. In terms of speech and facial expressions, enough work has been done (Sebe et al., 2004). Text-based emotion identification systems, on the other hand, still require more study. The recognition of human emotions in text is a challenge in computational linguistics. From a practical point of view, it is becoming increasingly crucial.

Emotion Detection will be important in the field of information technology & artificial intelligence, particularly in interfaces between humans and machines. To detect emotions using AI, different parameters must be taken into account. Various approaches are used to detect human emotions such as expressions on the face, movement of body, blood pressure, and heart rate, as well as text information. This research focuses on detecting emotions in text data. There are enormous number of textual materials available on the internet now a days. Extraction of emotion from diverse aims, such as from business is important (Douiji & Mousanif, 2015). For example, in the world of luxurious lifestyle, for purchasing items, the emotional aspect as a brand, distinctiveness, and reputations are far more important. In such conditions, the buyer can happily purchase the item even if the cost is high. Emotion detection may assist in marketing for a company's growth in customer relationship management, product creation, and service distribution.

To identify emotions from text there are four primary methods exists: 1) keyword-based 2) lexical affinity 3) learning-based 4) hybrid method approach. In all of the above methods there are issues with emotion recognition. The questions like what information should be used to extract features? How can we keep up with the ever-changing or evolving written terms utilized in ordinary conversations? Which types of emotion pointers may be found in a speech? How to extract contextual information? How may such traits be combined to get a desirable result? What emotions will be given to a text? Specially in the case of word combinations with several words. And which emotion categories should be utilized in training the data set? What is the finest machine learning classifier to use for wide-range text type data? And is it necessary to employ more than one classifier? There are issues and these issues need proper attention.
In this paper we design a system for recognizing emotion from text with great accuracy. We employ a supervised learning-based approach. When it comes to emotion detection, supervised text classification algorithms are typically work with quite high efficiency. However, in order for these strategies to be efficient, huge annotated data sets are necessary to create classification models. Therefore, we use supervised learning approach to detect the emotion.
In text-based emotion detection Natural Language processing (NLP) plays a vital role. NLP comprises of two parts: human language comprehension and human language generation. The Emotion Detection system's abstract flow is shown in Figure 1. The steps to follow are as follows: pre-processing of data, extraction of important features, and prediction model.

Figure 1

This paper includes different types of supervised learning methods of machine learning. As we know supervised learning works well with labelled data that means data tagged with its correct outputs. Different types of supervised classifier such as SVM, Decision tree Classifier, Logistic Regression, Naïve Bayes, Random Forest Classifier, KNN are used to inspect the accuracy, precision, recall and F1 score in emotion detection from textual data. When compared to unsupervised approaches, supervised learning-based methods can produce better outcomes. The purpose of this paper is to enable machines to recognize emotions in textual data and find out the efficiency of different classifiers on TBED.

The rest of the paper is organized as follows. In section 2, we review the literature. Section 3 explicitly explains several theories of emotion, the procedures of emotion detection, and different classifier of ML (Machine Learning). In section 4, we present and discuss our results. Section 5 concludes with future research plans.

LITERATURE REVIEW

Despite growing popularity, many people still use text to create connections with one another in everyday life and on social media. Text is still the most common way for people to convey their sentiments to other people. Today, the cyberspace is the key by which people share their feelings, and opinions (Gaind et al., 2018). Mining emotion from text is significantly more challenging due to the nature of the material. If it is possible to identify certain emotions in a text by looking at the words that define them then it is simpler to recognize emotions from the text.

Many academics from computer science, psychology, cognitive sciences, linguistics, and other fields have flocked to the topic of emotion detection. Facial expressions, voice, movements, and textual interactions are all ways for people to communicate their emotions. For detecting emotions, researchers all over the world prefer speech and face detection because they evoke emotion quite precisely and without any ambiguities. In comparison, there has been very little research done in emotion detection via text.

For emotion recognition, Herzig et al.(Herzig et al., 2017) employed a word vector that is previously trained. To boost the accuracy, they employed five separate datasets from various categories of emotion identification. The research scientist purpose ensemble approach achieves this using a linear model based on the BOW and a non-linear model based on pre-trained word vectors. Each text was displayed as a BOW using the BOW strategy and the SVM with a linear structure. They used the SVM with a non-linear structure to describe a text as a fixed-size vector that must be combined from word embedded vectors in the Word Embedding method. The results reveal that for domain-specific datasets, the ensemble technique out performs earlier strategies.

At the phrase level, Agrawal et al. (Agrawal & An, 2012) recommended an unsupervised, context-based technique to detect emotions from texts. Semantic and syntactic dependencies are used to classify the phrases. There is no requirement for a noted dataset or the use of a lexicon. Furthermore, by recommending the impacted semantic relatedness stages, addressing divergence modifiers, denials, and idioms, the system's efficiency is improved.

Multiple psychological categories are used by Hasan et al. (Hasan et al., 2019) to categorize emotions in different networks. Two classifications, hard and soft are accustomed group sentiments. To detect emotions throughout psychological classifications using hard classification labels, the Bayesian classifiers Naive Bayes and logistic regression are utilized. The research suggests that SVM might be applied for soft labelling. The authors worked upon a two-stage framework called EmotexStream to identify current twitter feeds for the second goal. The first step separates the tweets using a binary classification system that includes express emotion. The second step employs Emoted to do a fine-grained, specifically tweet emotional categorization.

On the other hand, Abdullah et al. (Abdullah et al., 2018) introduced the SEDAT model, that analyses the emotion of a statement in the Arabic language based on its amplitude. The model was evaluated using the SamEval 18 Task 1 dataset for the research. To obtain more accurate results and decrease fitting problem, they use a variety of word embedding approaches. Feedforward, CNN, and LSTM technologies make up the SEDAT. The results indicate that the designed model outperforms to the SVM model when the results of the two models are evaluated for analysis.

The emotive analysis of Roman Urdu is the subject of Ghulam et al. (Ghulam et al., 2019). To categorize the sentiments in the Roman Urdu language, the authors build a Deep Neural Network long short-term memory (LSTM). The results are assessed using baseline techniques such as SVM, RF, and Naive Bayes.

Kozareva et al. (Kozareva et al., 2007) performed emotion categorization using AlltheWeb, MyWay, and Yahoo data. They suggested that terms that appear often in a large number of texts about a specific emotion has the potential to communicate that feeling. Rage, terror, sorrow, amazement, hatred, and happiness are the emotions employed in this experiment. They analyzed the MIs (Mutual Information Scores) of a collection of substance and emotion terms by combining the intensity word counts from the three sites.

As an emotion lexicon, Tokuhisa et al. (Tokuhisa et al., 2008) inevitably created a set of 1.3 million emotional occurrences from the Japanese network. They selected 349 emotion terms from a Japanese dictionary and constructed a customized lexicon of sentiment words divided into 10 emotion types for generating an EP corpus. They used the emotion lexicon to collect 1.3 million occurrences as emotion origins from the Japanese online corpus, which comprises 500 million phrases. Then, with the help of an annotator, they analyzed 2000 incidents at random. Following that, they used a two-step procedure for emotion recognition that included sentiment polarization and classification of sentiments.

For automated TBED, Wang et al. (Wang et al., 2012) gathered one of the major sets of information from Twitter, around 2.5 million tweets. After pre-processing the data using seven emotion groups, they used two classifiers of ML (LIBLINEAR and MNB) to identify emotions. They also employed an attribute fusion of unigrams, bigrams, remaining emotion and POS to examine polarity using the LIWC dictionary and MPQA lexicon. Their designed system attains a maximum accuracy of 65.57 percent.

SVM are used to identify sentiments into seven categories on blogs by Aman et al. (Aman & Szpakowicz, 2008). The characteristics are corpus-based unigrams that include terms from Roget’s Thesaurus and WNA's emotion lexicons. In machine learning studies, they merged these characteristics and got better outcomes. On 21 observations, the combined label was chosen when more than one emotion or no emotion fit. For every class, F1 varied from 0.493 to 0.751.

In suicide notes data set, Cherry et al. (Cherry et al., 2012) described two techniques for text-based emotion detection (TBED). Their technique relies on a single classifier for each emotion, which makes label balancing and which is much easier to deal with rapid growth difficulties. It's a binary classifier resulting in better outcomes. They use Roget’s Thesaurus to come up with fifteen emotions which they employ in their model. They use a multi-label language classification called a latent sequence model to assign zero or more sentiments per each phrase. The F1 score is only 55.22%.

We see, there is gap in research. Emotion detection in texts uses fewer machine learning classifiers. To our knowledge only a small amount of work has been done in the area of fundamental emotion identification. In this paper, we develop a system that will capture the framework and serial structure of text which will enhance performance in sentiment recognition. Our goal is to compare the performance and accuracy of different classifiers.

THEORETICAL DEVELOPMENT/MODEL

Emotion detection is important in the development of human-machine interactions. Deep feelings directed towards somebody else in reaction to internally or externally occurrences are referred to as emotion. Emotion is a term that denotes a person's or a group's sentiments. As we have already discussed that emotion is a complex phenomenon and detecting emotion from textual data is a challenging task. Before going to different methods first we present different theories on emotion.

Studies on emotion categorization dates back to the Roman and Greek eras. In the nineteenth century, evolutionary emotions theories were developed by Charles Darwin and William James. Then, in 1991, Turkle proposed that computers are more than simply a mathematical tool; they can also be used to read emotions, languages, and civilizations. Scherer also created an intelligent system in the early 1990s that attempted to assign an emotional tag to the outcome of emotional experience data. Picard also examined emotional systems in human existence in 2003, presenting a notion of emotive computation. Because emotional intelligence is weak in computers, Picard encourages researchers to give machines the ability to recognize emotions in order to make the right decisions. Different researchers proposed different theories of emotion. Ekman (Ekman,1999) had created an emotion model in which the main parts were disgust, surprise, joy, sadness, and anger. For the use of facial emotion recognition, Ekman's model is quite beneficial. However, it is debatable if Ekman's model fully covers the whole range of emotions or it has limitations because it includes only Western civilizations' emotions. However, fear, sorrow, surprise, anger, love, and joy are among the six main emotions considered in Parrot's model. Parrott (Parrott, 2001), on the other hand, took this bigger classification and placed them in a tree like structure that eventually encompassed 100 different emotions. Russell et al. (Russell & Barrett, 1999) proposed a two-dimensional circumplex model with 150 emotional markers plotted along two orthogonal axes which is known as circumplex emotion model. The horizontal axis of arousal (pleasure) shows disagreeable in the negative, neutral in the middle, and delightful in the positive, whereas the vertical axis of arousal (activation) represents inactivation in the negative, neutrality in the core, and increased activity in the positive. Joy, enthusiasm, arousal, discomfort, dissatisfaction, discouragement, tiredness, and relaxation are all included in this circumplex paradigm (Cambria, 2016; Russell, 1980). The emotive aspects (anger, fear, guilt, jealousy, and so on) are defined by Russell et al. as parallel part of a circumplex and sheer as a fuzzy ladder (Russell & Barrett, 1999). They felt that feeling could be broken down into manageable chunks. Similar emotions, such as irritated, disturbed, and irritated, are put together in this approach, whereas different emotions are separated. Russell's paradigm unlike Ekman's, emphasizes emotional interactions. For example, when you're sad, you assume to have little or no happy sensations (Langroudi et al., 2018) Russell's emotion model employs to categorize emotions precisely on a margin of a circumplex type shape, using the $X^2 + Y^2 = B^2$ for each x and y coordinate. It does not, however, encompass the circle's feelings. To address this issue, Scherer (Scherer, 2005) devised a model in which equation $(xz)^2 + (ym)^2 = B^2$ the whole circle for each x and y feeling. As a result, the model encompasses emotions such as nil valence and a minor undesirable level of arousal. This model represents a wide variety of emotions and is 2D in terms of arousal and valence (Langroudi et al., 2018; Scherer, 2005). In order to determine every message's degree of ability and excitation, Perikos et al. (Perikos & Hatzilygeroudis, 2013) also utilized Sherer's approach for categorizing weblog material in 2D of sentiment (Perikos & Hatzilygeroudis, 2013).

In a set of phrases from the dictionary of affect, Whissell developed a continuing bi-dimensional sentiment model with both the axes of evaluation and activation. Whereas the activation factor includes taking any action that can be either passive or active, the evaluation factor includes detecting sentiments that can vary from negative to positive. Plutchik (Plutchik, 2001) postulated a three-dimensional combination model based on eight fundamental-composite feelings (sadness, joy, anger, trust, fear disgust, anticipation and surprise). Similar to a color wheel, he divided feelings into four groups: main, secondary, tertiary, and opposing emotions. The middle portion of the emotion wheel is where the basic emotions are located, while in the outside parts, the more complex emotions are situated. All these is further split into three emotional levels, the most intense of which is found closest to the inner circle. For instance, approval is less difficult than adoration but more complex than trust. Disgust is more nuanced than appreciation and less complicated than dislike and dullness. While opposite sentiments are divided by 180 degrees, similar feelings are clustered collectively. Ortony et al. (Ortony et al., 1988) proposed an OCC Emotion model with twenty-two sentiment classes derived from three foundations: goals, standards, and sensitivities, each of which is the foundation for three different kinds of schedules, actions, and objects. As a result, events indicate whether goals are desirable or undesirable, actions explain if standards are praiseworthy or blameworthy, and objects reflect whether people have attractive or disagreeable tastes. Furthermore, certain emotions are triggered by the combination of two or three other emotions. Steunebrink et al. gave a computer scientist's viewpoint on the OCC model, offering methods to make it more usable, practical, and computable (Ortony et al., 1988; Clore & Ortony, 2013). The Hourglass emotion model divides sentiments into 20 categories (semi good, semi bad) over four dimensions (Cambria et al., 2012). It is based on Plutchick's design. Each sentiment is represented by a couple of words that show the word's origin. According to Cambria et al. their suggested design can describe everyone's experienced emotion. The distinction among guilt and disgrace, for example, disgrace is the negative relationship between the self and the deed and Guilt is a sense of wrong doing brought on by the belief that if something wrong is done. When a person believes they are a horrible person, they experience shame. The distinction is significant because it shows distinct findings and accurately illustrates the similarities and differences between the effective words.

We see there are mainly four varieties of procedures for emotion detection such as **1)** Keyword based detection method 2) Lexical affinity method 3) Learning based method 4) Hybrid based methods. These methods are again subdivided into subcategories. Key word spotting method detects the word based on "word based", "line based" or "document based" approach. Lexical affinity method follows Lexican based approach. However, learning based method uses SVM, KNN, random forest, logistic regression and decision tree techniques. Hybrid based methods uses either of the two or three above methods. Our objective in this paper is to evaluate the performance of different models to see which model is more efficient with greater accuracy.

Procedure for Text Emotion Detection

We use the following steps to design our system. First, we create an input dataset using different databases. Next, using the 're' library we remove regular expressions and symbols and remove lemmas using the WordNetLemmatizer from NLTK. We also remove the multi-letter ambiguities e.g., 'noooo' is translated to 'no'. Then we tokenize data using tokenizer function. We remove the stop-words. Next, using the TF - IDF technique, we vectorize the texts. Following that we employ a supervised learning strategy. To train the system, we apply several machine learning classifiers like SVM, Nave Bayes, Logistic Regression, Decision Tree and Random Forest. Then we compare the results of different algorithms as well as their accuracies.

<u>Input Dataset</u>

To test our experiment, we created a stable dataset integrating Dailydialog, Isear, and Emotion-Stimulus with five markers: happy, sorrow, furious, awe, and inactive. Short messages and chat utterances constitute the majority of the texts. Table 1 summarizes the year, content, size and the emotion categories of the database. It also provides the information whether the dataset is a balanced dataset or not. As reported Dailydialog and Emotion-Stimulus are unbalanced and Isear is somewhat balanced dataset.

Table 1. Summary of the databases used in this study

Dataset	Year	Content	Size	Emotion categories	Balanced
Dailydialog	2017	Dialogues	102k sentences	happy, sorrow, furious, awe, inactive	No
Emotion-Stimulus	2015	Dialogues	2.5k sentences	happy, sorrow, furious, awe, inactive	No
Isear	1990	Emotional Situations	7.5k sentences	happy, sorrow, furious, awe, inactive	Yes

Table 2 reports the number of sentences across different emotion labels. For the emotion label “Happy”, “Sorrow”, “Furious”, “Inactive”, and “Awe” there are 2326, 2317, 2259, 2254 and 2171 sentences respectively. In total there are 11327 sentences used in the dataset. Before testing, we split the data set into two parts: one for training and another for testing**.** Size of the training and testing dataset equal to 7934 and 3393 sentences respectively.

Table 2. Emotion Label with Sentences:

Emotion Label	Number of Sentences
Happy	2326
Sorrow	2317
Furious	2259
Inactive	2254
Awe	2171

Preprocessing

Next, we pre-process the dataset using 1) Tokenization 2) Stop word removal 3) Normalization 4) Lemmatization 5) Stemming 6) POS tagging 7) Word embedding to make it noise free. For model efficiency and accurate processing time and for better accuracy, we clean unnecessary data from the dataset.

Feature Extraction

After completing the aforementioned processing procedures, the data are now error-free and prepared for extraction. To extract the features, we use TF-IDF i.e., Term frequency (TF) and Inverse Data Frequency (IDF) method. We convert the text sentences into numerical vectors using the TF-IDF technique.

- *Term Frequency (TF)* is defined as

$$tf_{t,d} = \frac{n_{t,d}}{Number\ of\ terms\ in\ the\ document}$$

Here the number n represents how many times the term "t" occurs in the text "d". As an outcome, a distinct TF value would be given to each document and word**.**

- *Inverse Data Frequency (IDF)* is defined as

$$idf_t = \ log\frac{Number\ of\ documents}{Number\ of\ documents\ with\ term\ 't'}$$

IDF determines the importance of a term.

- *Term frequency inverse document frequency* is defined as

$$(tf\_idf)_{t,d} = tf_{t,d} * idf_t$$

which measures how essential a term remains to a document in a dataset.

Training and Testing the Dataset

For training and testing the data, extracted features are chosen after feature extraction. The entire dataset is alienated into a training dataset and a testing dataset. The data which is used to train the model by providing the characteristics of numerous instances of an item is found in the training dataset. The testing dataset is then used to evaluate how successfully the model was trained using the training dataset. It is very challenging to test all training data. There are 100 percent data in the dataset and we take 30 percent from the dataset for testing to get better accuracy. The machine learning methods used for emotion detection are often classified as supervised classification. In this report we have used a supervised learning approach for detecting the emotion from the text. We have deployed different classifiers to classify the data. The classifiers used are as follows. 1) Naïve Bayes 2) Support Vector Machine 3) Logistic Regression 4) K Nearest neighbour 5) Decision Tree and 6) Random Forest Classifier. These classifiers are used to classify emotion from text data. Next, we use different performance measures such as accuracy, precision, recall and F1 score to measure the performance of each classifier. We also use Confusion Matrix to check the efficiency.

RESULTS AND DISCUSSION

We have applied six supervised classifiers such as Random Forest, Logistic Regression, Naïve Bayes, K Nearest Neighbour, Support Vector Machine and Decision Tree. Table 4 reports the results of different performance measures: accuracy, precision, recall and FI score of each classifier.

Table 4. Performance measures of different classifiers

SL #	Classifiers	Accuracy	Precision	Recall	F1 Score
1	Naïve Bayes Classifier	67.020	0.691	0.669	0.672
2	Random Forest Classifier	63.247	0.645	0.634	0.631
3	Logistic regression	69.348	0.695	0.694	0.693
4	Support Vector Machine	72.708	0.729	0.728	0.727
5	Decision Tree Classifier	54.936	0.553	0.552	0.551
6	K Nearest neighbor	44.149	0.617	0.448	0.442

We see that SVM has highest F1 score of 0.727. Next, Logistic regression has second highest F1 score of 0.693. KNN has the lowest F1 score of 0.442. The results also show that SVM has highest and KNN has the lowest accuracy**.** The graph below compares and present the results of different models.

Next, we use a confusion matrix to compare the results. The performance of classifiers on a number of test data for which the true values are known and is commonly described using a table termed as confusion matrix. Although the confusion matrix itself is rather easy to understand, but the terms used in connection with it can be confusing, so it is named as Confusion Matrix. To get better knowledge about the confusion matrix, let’s take a confusion matrix for binary classifiers that means the classifier produces only two outputs. In a confusion matrix, there are basically two values such as *Actual Value* and *Predicted Value*. Actual value is again divided into two subtypes such as *Actual Yes* and *Actual No*. In predicted value, it is also like *Predicted Yes* and *Predicted No*. The basic terms used in Confusion Matrix are defined as follows:

- True Positive (TP): When *Predicted Yes* is alike to *Actual Yes*.
- True Negative (TN): When *Predicted No* value is same as *Actual Yes*.
- False Positive (FP): When *Predicted Yes* is same as *Actual No*, we get *Type1 Error.*
- False Negative (FN): When *Predicted No* is same as *Actual Yes*, we get *Type2 Error.*
  We calculate the following variables for the Confusion Matrix.
- Accuracy = (TP+TN)/N (where N= total). It tests the correctness of classifier.
- Error Rate = (FP+FN)/N. It calculates the incorrectness of a classifier.
- Recall = TP/*Actual Yes***.** It calculates how much *Predicted Yes* is same as *Actual Yes*.
- Precision = TP/*Predicted Yes*. It calculates the correctness of *Predicted Yes*.
- Prevalence = *Actual Yes*/N (Where N=Total). It calculates the frequency of *Actual Yes*.
- F Score = It is the weighted average value of Recall and Precision.
- Specificity= TN/*Actual No*. It calculates how many times *Actual No* is same as *Predicted No*.
- False Positive Rate= FP/*Actual No*. How often *Actual No* equal as *Predicted Yes*.
  In this way confusion matrix works upon above mentioned different classifiers to achieve the efficiency of the classifiers.

Chart 1. Confusion Matrix of NAÏVE

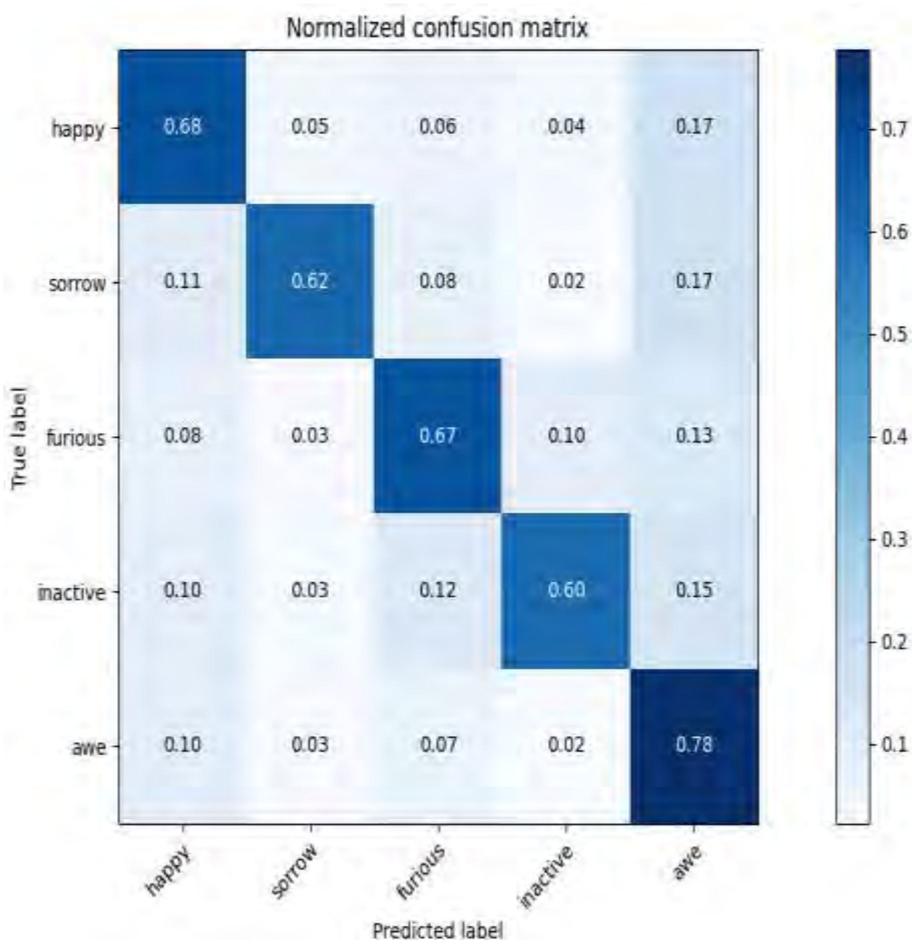


Chart 2. Confusion Matrix of Random Forest

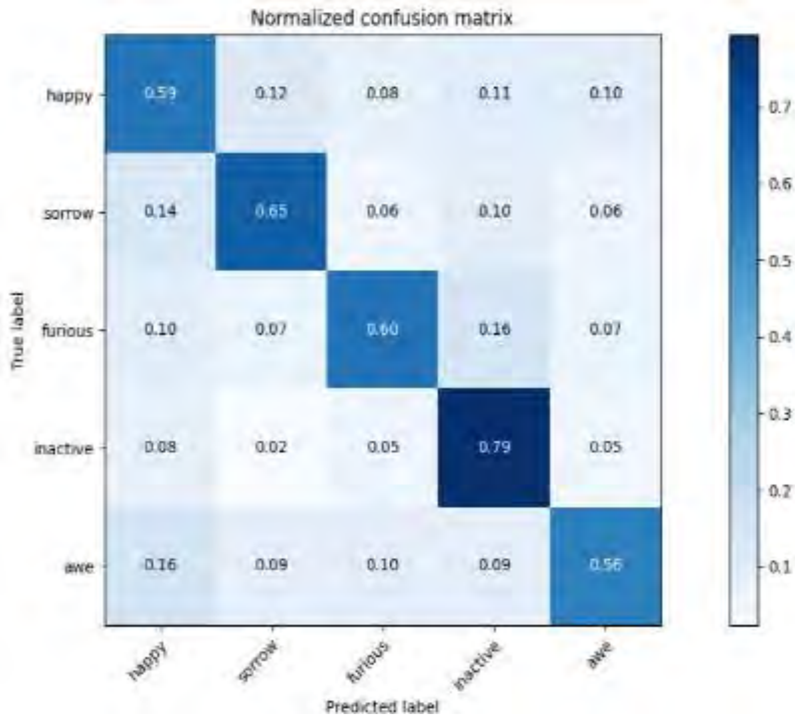


Chart 3. Confusion Matrix of Logistic regression

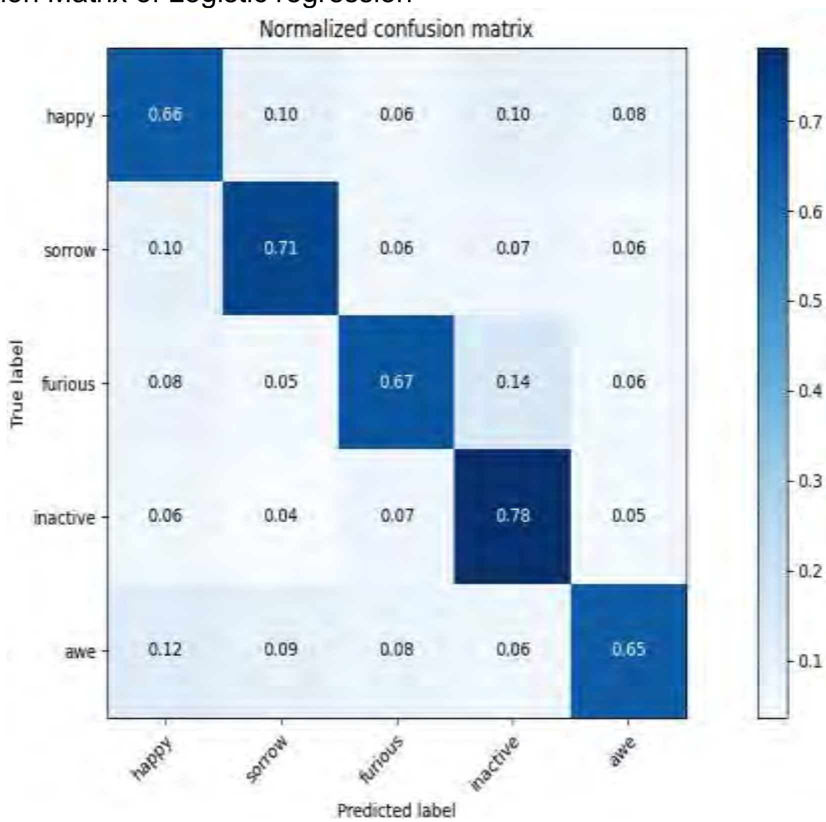


Chart 4. Confusion Matrix of SVM

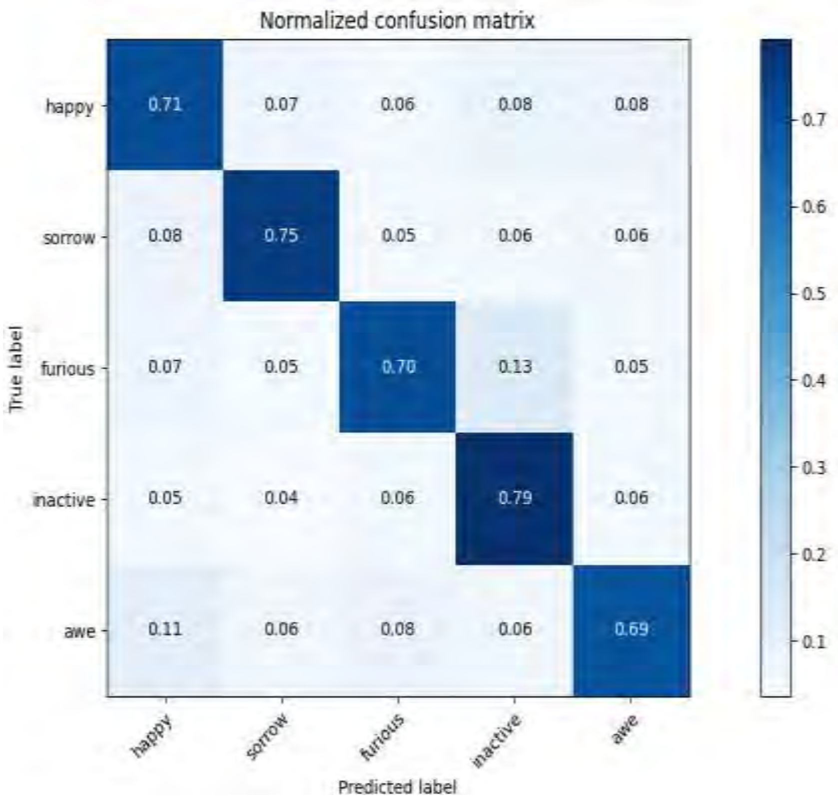


Chart 5. Confusion Matrix of Decision Tree

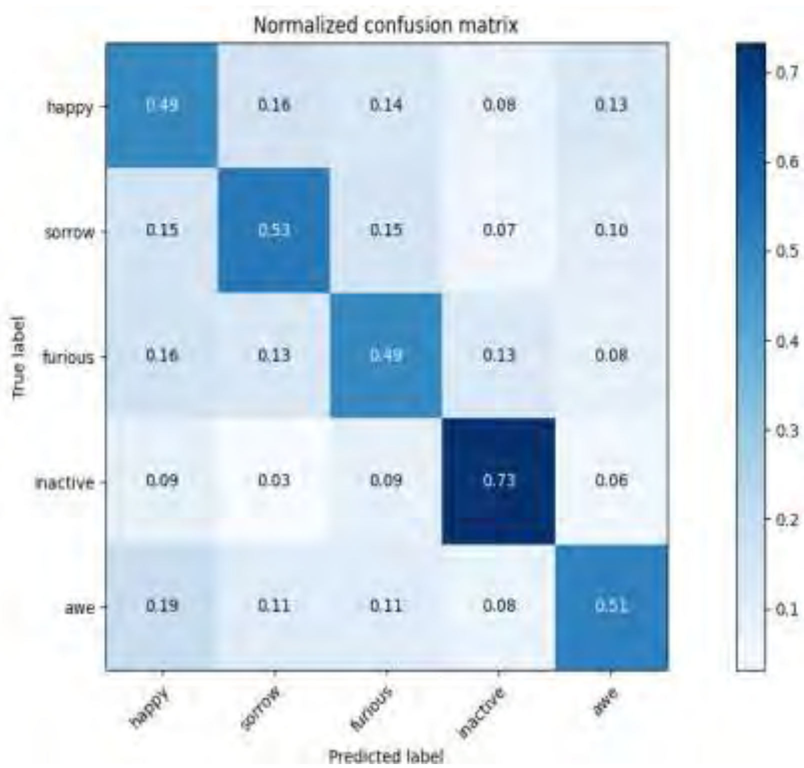
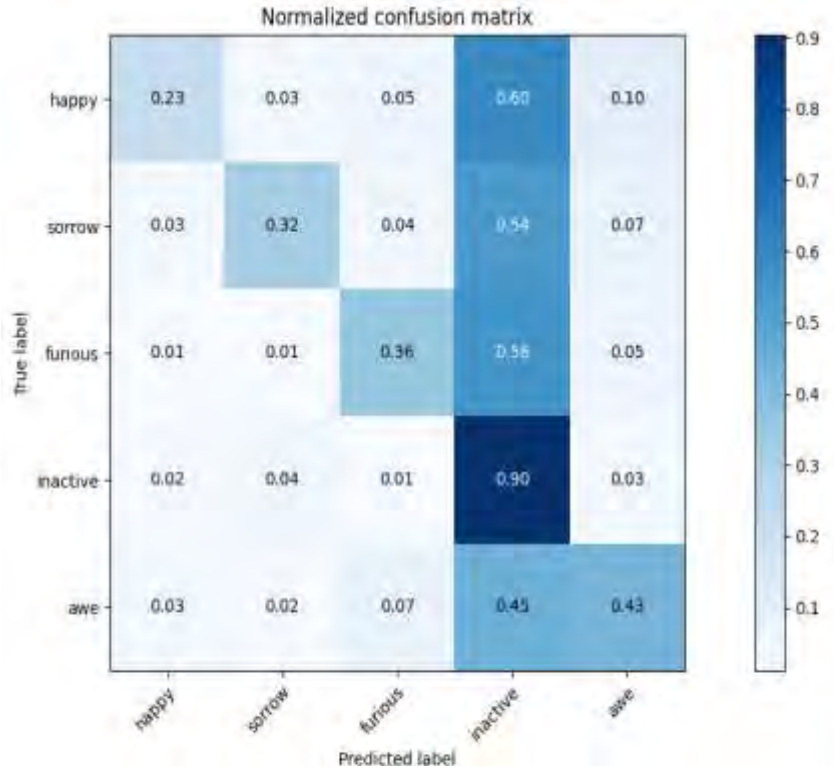


Chart 6. Confusion Matrix of KNN



After applying different classifiers, the above table and comparison table shows that the SVM has highest Precision, Accuracy, Recall and F1 score than other classifiers. The above table and figures indicate that SVM works better in emotion detection.

CONCLUSION AND FUTURE SCOPE

Though there are studies on emotion detection using facial expression, speech but very few works have done using emotion detection in text. We use more classifiers to produce the results. We see that SVM works more efficiently than other classifiers. Accuracy of SVM is 72.70%. F1score, recall and precision are 0.727,0.728 and 0.729 respectively on the task in detecting the emotion in text.

The main goal of this study is to identify the best technique for textual emotion detection. By creating a benchmark corpus and using various machine learning methods for emotion detection, the paper introduces a unique methodology for detecting emotions from text. However, this research is limited to only detecting “words”. It is even better to detect emotion in a sentence or paragraph. Finding a sentence's total impact may be useful in many areas, as marketing is one of them. Analyzing what consumers review might reveal their mental condition. More number of emotions can be recognized by adding additional features (emotions) to the system. It also seeks to demonstrate the connection between Valence, Arousal, and Dominance which can be implemented in online chatting platform to identify the emotion of a person.

REFERENCES

Abdullah, M., Hadzikadicy, M., & Shaikhz, S. (2018). SEDAT: Sentiment and Emotion Detection in Arabic Text Using CNN-LSTM Deep Learning, 17th IEEE International

Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, pp. 835-840, doi: 10.1109/ICMLA.2018.00134.

Agrawal, A., & An, A. (2012). Unsupervised emotion detection from text using semantic and syntactic relations, In Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society, 346–353. ISBN: 9780769548807

Aman, S., & Szpakowicz, S. (2008). Using Roget's Thesaurus for Fine-grained Emotion Recognition, In Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I, https://aclanthology.org/I08-1041

Binali, H., & Wu, C. (2010). Computational Approaches for Emotion Detection in Text, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.863.8593&rep=rep1&type=pdf , DOI: 10.1109/DEST.2010.5610650 · Source: IEEE Xplore

Cambria E. (2016). Affective computing and sentiment analysis, IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016, doi: 10.1109/MIS.2016.31.

Cambria, E., Livingstone, A., & Hussain, A. (2012). The Hourglass of Emotions. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R., Müller, V.C. (eds) Cognitive Behavioural Systems. Lecture Notes in Computer Science, vol 7403. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34584-5_11

Cherry, C., Mohammad, S. M., & De Bruijn, B. (2012). Binary Classifiers and Latent Sequence Models for Emotion Detection in Suicide Notes, Biomedical Informatics Insights, Vo; 5, pp. BII–S8933. doi:10.4137/BII.S8933

Clore, G. L., & Ortony, A. (2013). Psychological Construction in the OCC Model of Emotion. Emot Rev, 5(4):335-343. doi: 10.1177/1754073913489751. PMID: 25431620; PMCID: PMC4243519.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, E. N., Votsis, G., & Kollias, S. (2001). Emotion recognition in human-computer interaction, IEEE Signal Processing Magazine, Vol. 18(1), pp. 32-80, doi: 10.1109/79.911197

Douiji, Y., & Mousanif, H. (2015). I-CARE: Intelligent Context Aware system for Recognizing Emotions from text, 10th International Conference on Intelligent Systems: Theories and Applications (SITA), Rabat, Morocco, pp. 1-5, doi: 10.1109/SITA.2015.7358408.

Ekman, P. (1999). Basic emotions - Handbook of Cognition and Emotion. 98(45-60): 16, https://doi.org/10.1002/0470013494.ch3

Gaind, B., Syal, V., and Padgalwar, S. (2018). Emotion Detection and Analysis on social media, Global Journal of Engineering Science and Researches, ISSN 2348-8034, Pg. 78-89, https://doi.org/10.48550/arXiv.1901.08458

Garett, B. (2021). Brain and behaviour: an introduction to Behavioral Neuroscience, 6th ed.: SAGE Publications.

Ghulam, H., Zeng, F., Li, W., & Xiao Y. (2019). Deep Learning Based Sentiment Analysis for Roman Urdu Text, Procedia computer science, Volume 147, Pg, 131–135, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2019.01.202.

Hasan, M., Rundensteiner, E. & Agu, E. (2019). Automatic emotion detection in text streams by analyzing Twitter data. Int J Data Sci Anal 7, 35–51. https://doi.org/10.1007/s41060-018-0096-z

Herzig, J., Shmueli-Scheuer, M., & Konopnicki, D. (2017). Emotion Detection from Text via Ensemble Classification Using Word Embeddings. Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval.

Kozareva, Z., Navarro, B., Vázquez, S., & Montoyo A. (2007). UA-ZBSA: A Headline Emotion Classification through Web Information, In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 334–337, Prague, Czech Republic. Association for Computational Linguistics. https://aclanthology.org/S07-1072

Langroudi, G., Jourdanous, A., & Li, L. (2018). Music emotion capture: sonifying emotions in eeg data, In: Emotion Modelling and Detection in Social Media and Online Interaction symposium at the AISB 2018 Convention, 4-6 April 2018, Liverpool, UK. (KAR id:66564), https://kar.kent.ac.uk/66564/

Ortony, A., Gerald, L. C., & Collins, A. (1988). The Cognitive Structure of Emotions, Contemporary Sociology A Journal of Reviews, 18(6) Cambridge: Cambridge University Press.

Parrott, W. G. (Ed.). (2001). Emotions in social psychology: Essential readings. Psychology Press., https://psycnet.apa.org/record/2000-12576-000

Perikos, I., & Hatzilygeroudis, I. (2013). Recognizing Emotion Presence in Natural Language Sentences. In: Iliadis, L., Papadopoulos, H., Jayne, C. (eds) Engineering Applications of Neural Networks. EANN 2013. Communications in Computer and Information Science, vol 384. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41016-1_4

Plutchik, R. (2001). The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American Scientist, 89(4), 344–350. http://www.jstor.org/stable/27857503

Ramalingam V. V., Pandian A., Jaiswal A., & Bhatia N. (2018). Emotion detection from text, Journal of Physics: Conference Series 1000 012027 DOI 10.1088/1742-6596/1000/1/012027

Russell J. A., & Barrett L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant, Journal of personality and social psychology, vol. 76, no. 5, p. 805, PMID: 10353204 DOI: 10.1037//0022-3514.76.5.805

Russell J. A. (1980). A circumplex model of affect, Journal of personality and social psychology, vol. 39, no. 6, p. 1161 - 1178, 1980, https://doi.org/10.1037/h0077714

Scherer, K.R. (2005). What are emotions? And how can they be measured? Social Science Information, Vol. 44, Pg. 695 - 729.

Sebe, N., Cohen, I., Gevers, T., & Huang T. S. (2004). Multimodal Approaches for Emotion Recognition: A Survey, Proceedings of SPIE - The International Society for Optical Engineering 5670, DOI:10.1117/12.600746

Tokuhisa, R., Inui, K., & Matsumoto, Y. (2008). Emotion Classification Using Massive Examples Extracted from the Web, In Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008), pages 881–888, Manchester, UK. Coling 2008 Organizing Committee, https://aclanthology.org/C08-1111

Wang, W., Chen, L., Krishnaprasad, T., & Sheth, A. (2012). Harnessing Twitter 'Big Data' for Automatic Emotion Identification., International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing, Amsterdam, Netherlands, 2012, pp. 587-592, DOI: 10.1109/SocialCom-PASSAT.2012.119.

DECISION SCIENCES INSTITUTE

Environmental Policy and Corporate Default Risk: International Evidence

Md Ismail Haidar
University of Texas Rio Grande Valley
E-mail: mdismail.haidar01@utrgv.ed

Nam H. Nguyen
University of Texas Rio Grande Valley
E-mail: namhoang.nguyen@utrgv.edu

ABSTRACT

We study how environmental regulations affect firms' probability of defaulting on their debt obligations. Using the country-level environmental policy stringency index, which measures the degree of stringency of thirteen environmental policy instruments mainly related to climate and air pollution, we show that firms in countries with stricter environmental policies have lower default risk, which supports the idea that environmental regulations can induce innovation and efficiency improvements that enhance firms' competitiveness (Porter's hypothesis). Further evidence reveals that the effect is stronger for firms with sound financial performance and weaker for firms with strong external governance mechanisms and low environmental performance. The results remain robust to difference-in-differences estimation, entropy balancing, instrumental variables, alternative series, and an extensive set of control variables. Our study reveals important economic benefits of environmental policy stringency that have not been explored by previous research.

KEYWORDS: Environmental regulations, Default risk, Carbon tax, Difference-in-differences, Competitive advantage, External governance, Environmental performance

DECISION SCIENCES INSTITUTE

Excel-Based Continuous Review Inventory Game

Jay R. Brown
Loyola University Maryland
Email: jbrown11@loyola.edu

Maxim Bushuev
Morgan State University
Email: maxim.bushuev@morgan.edu

Tatiana Rudchenko
Georgia Institute of Technology
Email: tatianarudchenko@gatech.edu

ABSTRACT

This paper presents an Excel-based continuous review inventory management game designed for classroom use. The game provides an engaging and interactive way for students to learn about inventory management concepts. Students make decisions regarding order quantity and reorder point every simulated four weeks to minimize annual costs. Students can see the impact of their decisions on inventory levels and financial performance throughout the simulation. The game contains several products with different costs and demand patterns and is suitable for use in courses in operations management and supply chain management.

KEYWORDS: Inventory management, Simulation, Experiential learning, Classroom game

INTRODUCTION

Inventory management is a critical aspect of operations management, and it is crucial for organizations to maintain optimal inventory levels to ensure efficient operations and minimize costs. Classroom simulations and games have been recognized as effective tools for teaching inventory management concepts and providing students with hands-on experience in decision-making and problem-solving. In this research paper, we present an Excel-based continuous review inventory management game designed for classroom use, which allows students to practice inventory management concepts in a simulated business environment.

This Excel-based game is built upon the principles of the continuous review inventory system, which is widely used in practice. It features a user-friendly interface that allows students to input data, monitor inventory levels, and make decisions on order quantity and reorder point. The game provides immediate feedback on the impact of students' decisions and provides a final total cost that can be compared across the class. The game provides an interactive and engaging learning experience that allows students to learn about inventory management concepts such as order quantity, reorder point, safety stock, and lead time in a practical setting.

The purpose of this research paper is to introduce and evaluate the effectiveness of this Excel-based continuous review inventory management game as a teaching tool for courses in operations management and supply chain management. The paper will provide a detailed description of the game and its features, as well as a basic analysis of its impact on student learning. The paper will also discuss the game's potential for improving students' ability to apply inventory management concepts in real-world situations.

The effectiveness of simulation games and other interactive experiential learning tools as a pedagogical tool to enhance student engagement, critical thinking skills, and retention of course material has been well documented in the literature (Huang et al, 2022; Angell & O'Brien, 2019; Schmitt, 2017). Studies have shown that simulation games can improve students' understanding of inventory management concepts and provide them with a practical understanding of the challenges and trade-offs involved in managing inventory (Gökgöz & Erkip, 2018; Yen et al, 2016). Moreover, simulation games have been shown to be effective in improving students' decision-making skills and increasing their engagement and motivation in the learning process (Kim & Lee, 2014; Tavakkoli-Moghaddam et al, 2012).

In this paper, we describe our Excel-based continuous review inventory management game and present the results of initial classroom use to evaluate its effectiveness in teaching inventory management concepts. The results demonstrate the potential of the game as an effective tool for teaching inventory management concepts and provide insights into the factors that affect the effectiveness of simulation games in the classroom. This research paper contributes to the growing body of literature on the use of simulation games and other experiential learning tools in business education.

LITERATURE REVIEW

The literature review for this paper is divided into two sections. The first section explores the literature on the continuous review inventory model. The second section explores the literature on the use of games and simulation in business, and more specifically operations management, courses.

Literature Review of the Continuous Review Inventory Model

The continuous review inventory model is a widely used inventory management technique that aims to maintain optimal inventory levels by continuously monitoring and reviewing inventory levels. The reorder point is determined based on the lead time, the demand rate, and the safety stock level. The model ensures that inventory levels are maintained at an optimal level, preventing stockouts and excess inventory. The model also allows for timely replenishment of inventory, which can improve customer satisfaction and reduce costs associated with stockouts and overstocking. In this subsection, we examine the key features of the continuous review inventory model and explore its advantages and limitations.

The continuous review inventory model involves two variables, the order quantity and the reorder point. Harris (1913) developed the economic order quantity (EOQ) model, which helped to lay the foundation for the continuous review inventory model. Clark and Scarf (1960) introduced the concept of the reorder point and the role of safety stock to avoid the probability of stockouts and overstocking. Key formulas will be given in the Methodology section.

One of the advantages of the continuous review inventory model is its simplicity and ease of implementation. The model can be easily integrated into existing inventory management systems, and it does not require sophisticated mathematical models or complex calculations. The model is also flexible and can be adapted to different types of inventory, including raw materials, finished goods, and work-in-progress inventory.

Several studies have explored the effectiveness of the continuous review inventory model in improving inventory management. For example, Lee and Lim (2013) found that the continuous review inventory model was effective in reducing inventory levels and lead times in a manufacturing company. Similarly, Bao and Zhou (2015) found that implementing the continuous review inventory model in a retail company led to improved inventory accuracy and reduced stockouts.

However, the continuous review inventory model has some limitations. One limitation is determining the appropriate reorder point and safety stock level. The accuracy of the model is dependent on the accuracy of demand forecasts, which can be difficult to predict.

Literature Review of Classroom Simulations and Games

Simulation and games are increasingly used in business classes, and more specifically operations management classes, to enhance students' learning experience and improve their decision-making skills. This subsection provides an overview of the use of simulation and games in operations management classes and highlights their potential benefits.

Several studies have shown that simulation and games can effectively improve students' understanding of complex operations management concepts. For example, in a study by Tavakkoli-Moghaddam et al. (2012), students played a game designed to simulate a supply chain system in which students made decisions on production, inventory, and transportation. Students playing this game showed significantly higher levels of understanding compared to those who did not play the game. Similarly, in a study by Kim and Lee (2014), students who played a supply chain management game controlling a company called “Smart Gear” showed higher levels of knowledge retention compared to those who did not play the game.

Simulation and games have also been shown to enhance students' decision-making skills. For example, in a study by Joglekar et al. (2005), students who played a simulation game on quality management showed significant improvements in their decision-making skills. Similarly, in a study by Chen et al. (2014), students who played a project management simulation game showed significant improvements in their decision-making skills related to project management.

As these references demonstrate, simulation and games have the potential to enhance students' learning experience and improve their decision-making skills in operations management classes.

METHODOLOGY

This section will discuss the game in three sections entitled Game Design, Game Overview, and Presenting to Students.

Game Design

The continuous review inventory game was created to give students hands-on experiential exposure to the inventory concepts of the continuous review model. Students will manage the chosen pre-populated product. Different products are set up with different demand patterns. Some products have low variability and stable demand while others have high variability, varying variability, varying demand rates, or some combination of these. In addition, new products can be created to suit the instructor’s needs.

Excel was chosen as the platform for the simulation because of its ubiquity and ease of creation. Students can see all the relevant parameters such as demand rate, demand standard deviation, order cost, carrying cost, lead time, stockout cost, current inventory, and incurred costs throughout the game. The game involves macros to move the game forward in 4-week increments. Students manage the chosen product for a full year this way.

The main objective of the game is to minimize inventory costs. To achieve low costs, students can calculate the EOQ and use this to guide the order quantity, *Q*, decision. Equation 1 gives the classic EOQ formula.

$$\boldsymbol{EOQ} = \sqrt{\frac{\boldsymbol{2DS}}{\boldsymbol{H}}} \quad (1)$$

where
D = demand rate per period;
S = order cost per order;
H = holding or carrying cost per period.

For the reorder point, *R*, students can either calculate this reorder point based on a desired service level or they may play the game a few times and naturally settle on an appropriate *R* by observing the inventory levels and adjusting as they go. This can be discussed with the students during the debrief. Equation 2 gives the classic reorder point, *R*, formula.

$$\boldsymbol{R} = \boldsymbol{dL} + \boldsymbol{z\sigma_d}\sqrt{\boldsymbol{L}} \quad (2)$$

where
d = demand rate per unit time
$\boldsymbol{\sigma_d}$ = standard deviation of *d*
L = lead time in the same units as *d*
z = z-score based on desired service level

The next section entitled Game Overview will show the layout of the game and discuss how it works.

Game Overview

Harvard Publishing has a simulation for periodic review called Inventory Basics. In this game, students manage the inventory of one of three products (wrench, paint, or rock salt) at a hardware store on a weekly basis. Each week you decide how many units to order if you decide to order at all. The continuous review inventory game, on the other hand, involves setting the order quantity, Q, and the reorder point, R, and then 28 days of demand are generated. Students manage the product through 13 periods, which covers a full year of demand.

To play the game, students open the Excel file and enable macros. Instructions are provided for students that may run into macro security issues. The instructor shows a slide that describes how to enable macros for those that have issues for both PC and Mac.

The Excel spreadsheet shows students just one sheet with all the relevant information in one screen. Figure 1 shows the view of the entire game. Each section of the spreadsheet will be shown in greater detail and explained.

Figure 1: Game view

Product	Hammers	
Avg Demand	0	per day
Std Dev of Demand	0	per day
Order Cost	$ 40.00	per order
Carrying Cost	$ 0.04	per day
Lead Time	3.50	days
Std Dev of Lead Time	0.29	days
Stockout cost	$ 25.00	per unit
Starting inventory	125	units

Planning Period: 14

User ID (before @ .edu)

Begin or reset

Order Quantity 200

Reorder Point 79

Advance to next period

TOTAL COST $ 3,064.12

Per 100 $42.03

Simulation Complete

COPY DATA

ENTER USER ID!

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	
Starting Inv	125	165	-1	44	81	118	152	195	36	72	106	154	188	
Ending Inv	165	-1	44	81	118	152	195	36	72	106	154	188	35	TOTAL
Total Demand	560	566	555	563	563	566	557	559	564	566	552	566	553	7290
Order Cost	$ 120	$ 120	$ 120	$ 80	$ 120	$ 120	$ 120	$ 120	$ 120	$ 80	$ 120	$ 120	$ 120	$ 1,480
Carrying Cost	$ 105	$ 97	$ 109	$ 104	$ 102	$ 116	$ 106	$ 118	$ 124	$ 115	$ 113	$ 107	$ 117	$ 1,434
Stockout Cost	$ -	$ 25	$ -	$ 50	$ 25	$ -	$ 50	$ -	$ -	$ -	$ -	$ -	$ -	$ 150
Total Cost	$ 225	$ 242	$ 229	$ 234	$ 247	$ 236	$ 276	$ 238	$ 244	$ 195	$ 233	$ 227	$ 237	$ 3,064

Q setting	200	200	200	200	200	200	200	200	200	200	200	200	200
R setting	79	79	79	79	79	79	79	79	79	79	79	79	79

Period 13

Day	Begin Inv	Demand	End Inv	Order Made	Order Rec	Stockout Cost	Order Cost	Carrying Cost
336	8	20	188		200			$ 7.52
337	188	20	168					$ 6.72
338	168	19	149					$ 5.96
339	149	20	129					$ 5.16
340	129	20	109					$ 4.36
341	109	19	90					$ 3.60
342	90	20	70	200			40	$ 2.80
343	70	18	52					$ 2.08
344	52	19	33					$ 1.32
345	33	18	15					$ 0.60
346	15	20	195		200			$ 7.80
347	195	21	174					$ 6.96
348	174	18	156					$ 6.24
349	156	19	137					$ 5.48
350	137	21	116					$ 4.64
351	116	20	96					$ 3.84
352	96	19	77	200			40	$ 3.08
353	77	21	56					$ 2.24
354	56	18	38					$ 1.52
355	38	20	18					$ 0.72
356	18	21	197		200			$ 7.88
357	197	19	178					$ 7.12
358	178	20	158					$ 6.32
359	158	20	138					$ 5.52
360	138	21	117					$ 4.68
361	117	20	97					$ 3.88
362	97	22	75	200			40	$ 3.00
363	75	20	55					$ 2.20
364	55	20	35					$ 1.40

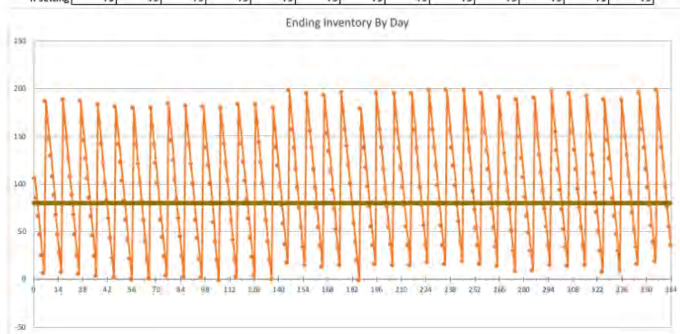


Figure 2 zooms in on the upper left section. This section shows all the relevant parameters of the game. Students can use these to decide on order quantity and reorder point. In the pilot test of this game, undergraduate students were told lead time was a constant 4 days. Along with more advanced products, the graduate version of this game changes lead time to be variable as Figure 2 shows to introduce more complexity. The only cells that students can interact with are colored yellow. There is a drop-down for students to select the assigned product. Here “Hammers” is selected. Selecting another product changes the associated parameters. Three yellow cells allow typing. The “Order Quantity” and “Reorder Point” cells are where students enter their desired values. The yellow cell below “User ID” is where students enter their user ID. This is used for homework. Three buttons are used in the game. The “Begin or reset” button clears any data and returns the game to the beginning. The “Advance to the next period” button saves the students values for Q and R and advances the game to the next period. This also updates the cost table and inventory graph, which will also be detailed in this section. The “Copy Data” button is used for homework. This button copies a string of data that records how well the student performed. It contains information on order quantities, reorder points, inventory costs, total demand, total cost, and some validation elements so students cannot change the string values without violating the validation. In future instances of this game, students will paste these strings into a learning platform quiz. The instructor can then download the responses into a spreadsheet to validate the strings and assign points according to a table that translates total cost for each assigned product into points for that product.

Figure 2: Close-up of parameters, buttons, and entry fields

Product	Hammers	
Avg Demand	20	per day
Std Dev of Demand	1	per day
Order Cost	$ 40.00	per order
Carrying Cost	$ 0.04	per day
Lead Time	3.50	days
Std Dev of Lead Time	0.29	days
Stockout cost	$ 25.00	per unit
Starting inventory	125	units

Planning Period: 1

User ID (before @ edu)

Begin or reset

Order Quantity 200

Reorder Point 79

Advance to next period

TOTAL COST $ -

Per 100

COPY DATA

ENTER USER ID!

Figure 3 shows the inventory by day graph, which updates each time the student advances the game to the next period. Students can use this to quickly see how well their reorder point (set here as 79 throughout the entire game) is performing.

Figure 3: Close-up of inventory by day graph

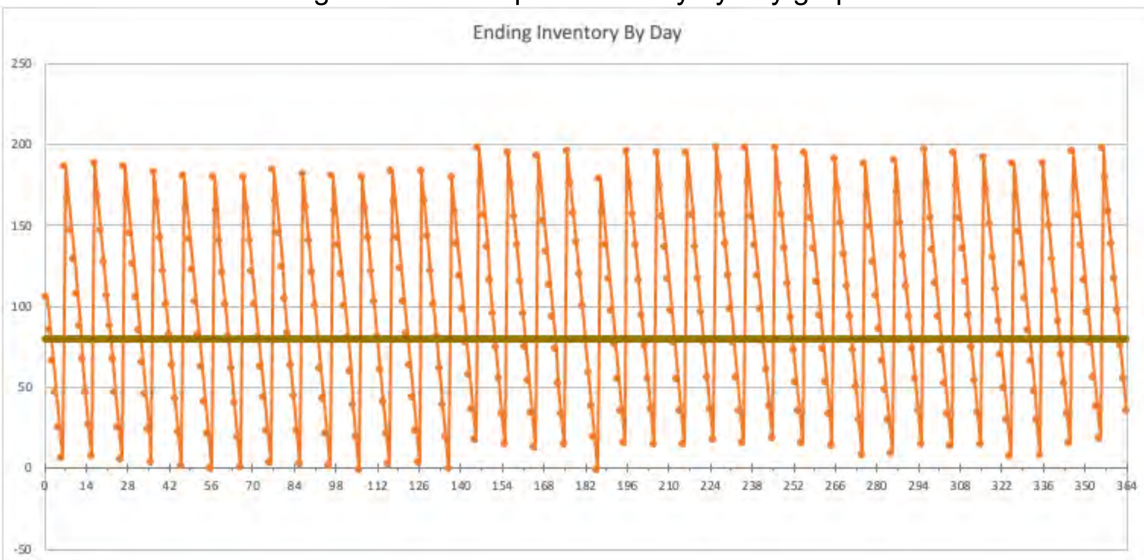


Figure 4 shows a close-up of daily demand, corresponding inventory on-hand, order details, order cost incurred, and carrying cost incurred. If inventory goes negative, these are considered back-ordered units.

Figure 4: Close-up of inventory details by day

Period 13

Day	Begin Inv	Demand	End Inv	Order Made	Order Rec	Stockout Cost	Order Cost	Carrying Cost
336	8	20	188		200			$ 7.52
337	188	20	168					$ 6.72
338	168	19	149					$ 5.96
339	149	20	129					$ 5.16
340	129	20	109					$ 4.36
341	109	19	90					$ 3.60
342	90	20	70	200			40	$ 2.80
343	70	18	52					$ 2.08
344	52	19	33					$ 1.32
345	33	18	15					$ 0.60
346	15	20	195		200			$ 7.80
347	195	21	174					$ 6.96
348	174	18	156					$ 6.24
349	156	19	137					$ 5.48
350	137	21	116					$ 4.64
351	116	20	96					$ 3.84
352	96	19	77	200			40	$ 3.08
353	77	21	56					$ 2.24
354	56	18	38					$ 1.52
355	38	20	18					$ 0.72
356	18	21	197		200			$ 7.88
357	197	19	178					$ 7.12
358	178	20	158					$ 6.32
359	158	20	138					$ 5.52
360	138	21	117					$ 4.68
361	117	20	97					$ 3.88
362	97	22	75	200			40	$ 3.00
363	75	20	55					$ 2.20
364	55	20	35					$ 1.40

Figure 5 shows a summary of all the relevant metrics by period. Here students can see the financial impact of their decisions on order cost, carrying cost, stockout cost, and total cost. This table populates as the game progresses.

Figure 5: Close-up of results for each 4-week period

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	
Starting Inv	125	165	-1	44	81	118	152	195	36	72	106	154	188	
Ending Inv	165	-1	44	81	118	152	195	36	72	106	154	188	35	TOTAL
Total Demand	560	566	555	563	563	566	557	559	564	566	552	566	553	7290
Order Cost	$ 120	$ 120	$ 120	$ 80	$ 120	$ 120	$ 120	$ 120	$ 120	$ 80	$ 120	$ 120	$ 120	$ 1,480
Carrying Cost	$ 105	$ 97	$ 109	$ 104	$ 102	$ 116	$ 106	$ 118	$ 124	$ 115	$ 113	$ 107	$ 117	$ 1,434
Stockout Cost	$ -	$ 25	$ -	$ 50	$ 25	$ -	$ 50	$ -	$ -	$ -	$ -	$ -	$ -	$ 150
Total Cost	$ 225	$ 242	$ 229	$ 234	$ 247	$ 236	$ 276	$ 238	$ 244	$ 195	$ 233	$ 227	$ 237	$ 3,064
Q setting	200	200	200	200	200	200	200	200	200	200	200	200	200	
R setting	79	79	79	79	79	79	79	79	79	79	79	79	79	

Figure 6 shows how this section changes when a product is chosen that has demand that varies by period. Notice the addition of the orange text that shows the seasonal forecast. In addition, the parameters section changes to align with the forecast for the appropriate period. In this way, students can create formulas in Excel to set *Q* and *R* and these will change as the period changes to correspond to the current period's demand and standard deviation.

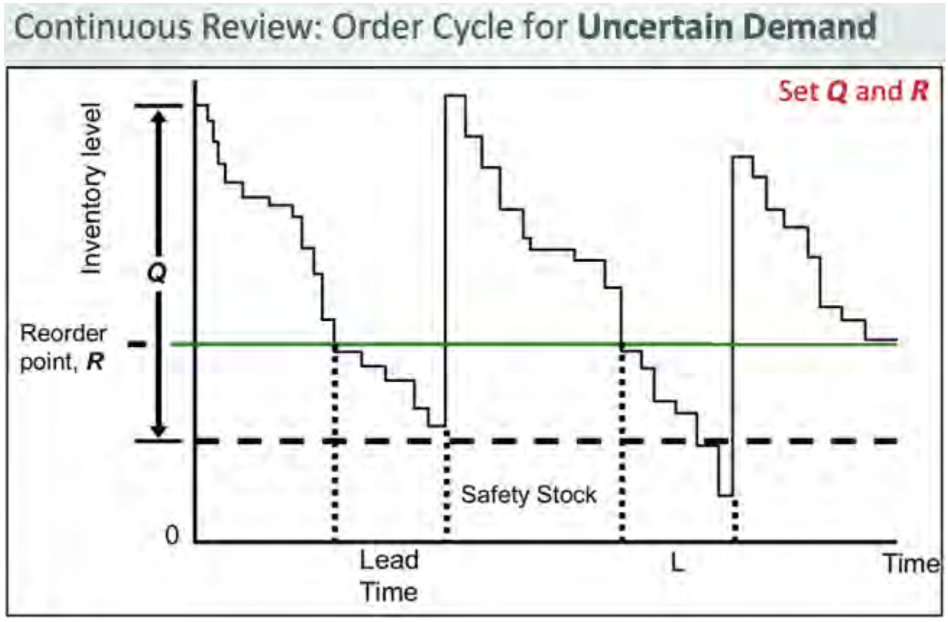
Figure 6: Close-up of results section with a product demand that varies by period

Period	1	2	3	4	5	6	7	8	9	10	11	12	13	
Starting Inv	125	44	136	46	139	121								
Ending Inv	44	136	46	139	121	102								TOTAL
Total Demand	439	445	448	444	675	676								3127
Order Cost	$ 120	$ 80	$ 120	$ 80	$ 120	$ 120								$ 640
Carrying Cost	$ 101	$ 100	$ 99	$ 104	$ 129	$ 127								$ 661
Stockout Cost	$ -	$ -	$ -	$ -	$ -	$ -								$ -
Total Cost	$ 221	$ 180	$ 219	$ 184	$ 249	$ 247								$ 1,301
Forecast	16	16	16	16	24	24	32	32	40	40	40	24	24	
Forecast Std Dev	1	1	1	1	1.5	1.5	2	2	2.5	2.5	2.5	1.5	1.5	
Q setting	179	179	179	179	219	219								
R setting	66	66	66	66	98	98								

Presenting to Students

In the initial test of the game, students in an undergraduate operations and supply chain management class played the game in the class session after the first inventory management session, which included the EOQ calculation. They were shown a slide (Figure 7), which explained an overview of the continuous review model and the game was explained and demonstrated. After the students opened the spreadsheet and all issues with enabling macros were overcome, they were told to select "Hammers" and play as many times as they wanted and record total cost after each attempt. For this class, lead time, *L*, was shown as a constant 4 days. In addition, a message would pop up if they set *R* to less than or equal to 70, which is below expected demand during lead time, to inform them of the implications of that setting.

Figure 7: Slide presented prior to introducing the continuous review game



After a few minutes, a PollEverywhere poll was presented asking them to submit the lowest score they were able to achieve. Following this, students were introduced to the reorder point formula along with examples and a poll problem for them to calculate and submit. For this reorder point calculation, 92% (45/49) of students submitted the correct response. Next students were reminded of the EOQ formula and again shown the continuous review model. At this point, students were told to play the game again and record total cost after each attempt.

RESULTS

In the initial testing phase of the game as described previously, undergraduate students were asked to play the game and record their total cost after each attempt. In the poll results, ten students needed to be removed based on infeasible answers. For example, a response of $1300 is impossible to achieve for the "Hammers" product. Of the remaining 36 students that submitted feasible responses to the first attempt and second attempt poll, the average first attempt lowest total cost was $4161.70 with a median total cost of $3987.50. In this first attempt, 13.9% were able to achieve a total cost below $3100. During the debrief (after the second attempt), students are told that recording a total cost at or below $3100 is easily achievable when setting optimal values of *Q* and *R* based on (1) and (2).

In the second attempt, after learning the reorder point formula and being reminded of the EOQ formula from the last class, the average total cost fell to $3168.10 with a median total cost of $3006.50. During this attempt, 83.3% were able to achieve a total cost at or below $3100. 86.1% of students showed improvement during attempt 2 while 11.1% stayed the same, and 2.8% recorded a worse score. However, the one student with the higher total cost was only by a few dollars and both attempts were below the $3100 threshold. The four students that stayed the same achieved a total cost at or below $3100 for both attempts. Table 1 summarizes the results from this initial test.

Table 1: Summary of Results

METRIC	ATTEMPT 1	ATTEMPT 2
Number of students	36	36
Average total cost	$4161.70	$3168.10

Median total cost	$3987.50	$3006.50
$3100 or less	13.9% (5)	83.3% (30)

The five students that scored at or below $3100 during their first attempt had something in common. They each used the inventory graph (Figure 3) to intuitively settle on an optimal or near-optimal reorder point through trial and error. By observing stock-outs or too much inventory on-hand when the replenishment arrived, they were able to adjust their reorder point appropriately.

DISCUSSION AND CONCLUSIONS

The Excel-based continuous review inventory game is an experiential learning tool to allow students to apply inventory theory concepts in a simulated real-world environment. The findings of this research underscore the value of incorporating such game-based approaches to operations management and more specifically inventory management education. As educators continue to explore innovative pedagogical methods, this game has demonstrated its potential as a tool for increasing student engagement and applying inventory management concepts.

Future enhancements to this game will allow it to be used in more ways. The game will have an option to be played in a way to teach the newsvendor or single-period model, which was first introduced by Scarff (1959). In this mode, the game will not include a reorder point and the student must decide how much inventory to order to cover 28 days. Any leftover inventory at this point will be scrapped for the pre-defined salvage value. A currently functioning extension allows the game to be set to include variable lead time, which adds complexity and a different reorder point formula. In addition, more complex products with varying demand by period can simulate seasonality and/or a trend in demand. Finally, highly variable demand can increase the challenge for graduate students or more advanced undergraduate courses.

The Excel-based continuous review inventory game has made a significant contribution to the field of inventory management education. The game created an interactive and immersive learning experience that enhanced students' understanding of inventory management principles and their ability to apply these concepts in real-world scenarios. Students were engaged in the game and showed considerable improvement in performance after learning about the reorder point formula. The game allowed them to immediately put this new knowledge to use in a simulated real-world environment. The game demonstrated the ability to enhance student’s understanding of key inventory management concepts, proved effective in improving students’ decision-making, and increased students’ engagement in inventory management material.

REFERENCES

Bao, X., & Zhou, M. (2015). Research on the Application of Continuous Review Inventory Model in Retail Enterprises. *2015 International Conference on Education, Management, Information and Medicine* (145-148). Atlantis Press.

Chen, J. C., Huang, Y. H., & Wu, Y. H. (2014). Improving students' decision-making skills in project management through simulation-based learning. *Journal of Educational Technology & Society*, 17(2), 1-13.

Clark, L. A., & Scarf, H. (1960). Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4), 475-490.

Gökgöz, F., & Erkip, N. (2018). Using simulation games in inventory management education: An empirical study. *International Journal of Industrial Ergonomics*, 68, 153-160.

Harris, F. W. (1913). An Approximation Solution for the Economic Lot Size Model with a Constrained Capacity. *Journal of Operations Research Society of America*, 1(1), 22-27.

Huang, Y. M., Silitonga, L. M., & Wu, T. T. (2022). Applying a business simulation game in a flipped classroom to enhance engagement, learning achievement, and higher-order thinking skills. *Computers & Education*, 183, 104494.

Joglekar, N. R., Moore, J. C., & Sayood, A. (2005). Developing decision-making skills in quality management: a simulation-based approach. *Decision Sciences Journal of Innovative Education*, 3(2), 139-160.

Kim, J., & Lee, J. (2014). Learning by gaming: Evaluation of an experiential simulation game for teaching inventory management. *Computers & Education*, 78, 101-109.

Kim, Y. H., & Lee, K. (2014). Effect of a supply chain simulation game on students' knowledge retention. *Journal of Business Research*, 67(5), 885-890.

Lee, S.-H., & Lim, J.-Y. (2013). Application of continuous review inventory model to reduce lead time and inventory level. *Journal of the Korean Society for Precision Engineering*, 30(7), 659-664.

Scarff, R. C. (1959). The optimization of a compound Poisson process. *Naval Research Logistics Quarterly*, 6(1), 47-54.

Tavakkoli-Moghaddam, R., Haji, R., & Sepehri, M. M. (2012). Evaluating the effectiveness of a simulation game on supply chain management education. *Computers & Education*, 58(1), 112-122.

Tavakkoli-Moghaddam, R., Makui, A., & Razmi, J. (2012). Designing a simulation game for inventory management education: A supply chain management approach. *Computers & Education*, 58(4), 1234-1244.

Yen, C.-L., Chiu, Y.-F., & Chuang, Y.-C. (2016). Development of an educational game for inventory management education. *Computers & Education*, 99, 56-66.

DECISION SCIENCES INSTITUTE
Explore the employee-organization relationship

Pei-Chen Chen
Department of Business Administration, Tainan University of Technology
Email: pcchenedu@gmail.com

Ming-Chao Wang
College of Management, Yuan Ze University
Email: mcwangedu@gmail.com

ABSTRACT

When flexibilization and fragmentation become trends, the employee-organization relationship (EOR) should be applied in dealing with multi-employment. Based on EOR, we agree that the organization is a critical player with a hands-on role in EOR. Therefore, an organization should pay attention to employee concerns to enhance the organization's benefits. Agency theory defines and prescribes a set of relationships in an organization. Thus, this research expounds on the EOR by using agency theory, discusses how extrinsic and intrinsic motivation affects an organization's agency problems and cost, and further completes the moderated mediation model of the EOR on an empirical basis.

KEYWORDS: Extrinsic motivation, Intrinsic motivation, Employee-organization relationship

INTRODUCTION

In the emerging and theoretical literature, there is a consensus that employees can provide a unique source of competitive advantage. If properly configured, organizational human resource policies can offer a direct and economically significant contribution to organizational performance (Huselid, 1995). The impact of organizations combining internal and external labor markets, specifically when they are more likely to fill jobs by external employment, is essential for understanding modern employment systems and governance, resulting in numerous studies comparing traditional and new employment trends (Kuvaas & Dysvik, 2010). Following this situation, it is critical in organization governance to determine how to manage employees with different managerial agents under this complex condition. Tsui et al. (1995) state an EOR strategy based on the employer's expectations about specific contributions.

In a social exchange relationship, one party (the organization) voluntarily provides a benefit to another (the employees), involving an obligation of the other party to reciprocate by giving some advantage in return (Whitener et al., 1998; Zhang et al., 2008). Based on Tsui et al. (1995; 1997), social exchange is the mechanism linking the EOR and organization outcomes. It desires effort on the part of employees and uses inducements to affect the expected contributions. Drawing on the social exchange on EOR, Tusi et al. (1995) assume that the organization may adopt different job employments to achieve maximum flexibility.

Agency theory is one of the richest theoretical traditions in management literature. The initial formulation of agency theory simplifies the relationship between the employees (agents) and the organization (principal), especially in solving goal conflicts and risk sharing of costs (Eisenhardt, 1989; Wiseman et al., 2012). Agency theory focuses on the relationship between the principal (employer) and agent (employee) via contracts. EOR management, however, should not only focus on exploring what EOR is and discussing the importance of EOR but also emphasize how to manage EOR. Few studies examine optimal governance efficiency in an EOR context (Tsui & Wu, 2005; Tsui et al., 1997). Therefore, in this study, we develop a framework that provides a general theoretical rationale for explaining how to manage the EOR from the social perspective of agency theory.

Social agency concepts (Wiseman et al., 2012) offer insights on how to achieve congruence between the employment relationship intended by the employer and the psychological contracts perceived by the employee. Motivation provides relevant concepts for driving employee efforts to aid organizational performance.

Motivation refers to the set of processes that encourages, directs, and maintains personal behavior toward attaining the goals of organizations (Greenberg & Baron, 2000). Under multi-employment, goals and information opacity vary by according to the institutional context. Individuals will differ in their goals, and these differences can translate into agency costs for principals. For example, compared to regular employees, non-regular employees are short-term and less insecure, so form of employment, the organization finds it difficult to monitor them, leads leading to higher agency costs. In our research, in structuring a moderated mediation model, we consider motivation as driving employees' actions to structure the moderated mediation model. This study focuses mainly on the relationship between the user company and agency workers. In a triangular employment relationship, the user company (the principal) is not only the supplier of labor conditions but also the decision-maker who offers a stable job to agency workers. Thus, this study seeks to take the agency theory as the basis for the theoretical framework. Under this framework, this study attempts to explore possible agency problems and agency costs between the user company and agency workers and examine whether the promotion-focused regulatory system will moderate the relationship between agency problem and agency cost.

LITERATURE REVIEW

Agency Theory

Agency theory, which was proposed in economics in the 1960s (Jensen & Meckling, 1976), deals with the different conflicts resulting from self-interest and opportunism between principals and agents. Based on the agency theory, organizational assumptions include information asymmetries, goal conflicts between principals, and between principals and agents (Filatotchev & Wright, 2011).

Jensen and Meckling (1976) view agency relationship as a form of contract which contains regulations giving the agent certain powers. However, the positions of the principal and the agent differ, and due to differences in personal benefit, conflicts in interest occur. Thus, agency problems result (Fama, 1980; Eisenhardt, 1989).

That is to say, agency theory assumes that both the principal and the agent are rational, the owner of the economic resources is the principal (banking management), and the employees charged with controlling and using these resources (both full-time employees and agency employees) are the agents. In the process of agreeing to agency contracts, both parties attempt to pursue the greatest possible self-interest or wealth accrual. Under this assumption, in order to avoid incorrect behavior on the part of employees, the principal will attempt to find suitable incentives and supervision measures in order to restrain the agents. In order to gain the trust of the principal, agents will restrain their own behavior. This tendency is defined as agency cost (Eisenhardt, 1989).

A major reason for agency problems is asymmetrical information and goal inconsistency (Bergen et al., 1992; Eisenhardt, 1989). When the principal and the agent have different goals and both have asymmetrical information, the agent will adopt the behavior or decisions beneficial to himself instead of to the principal; therefore, the principal incurs losses and agency problems arise (Eisenhardt, 1989; Fama, 1980). The agency relationship exists when the people the employment agency hired are working for the user company; subsequently, agency problems result (Ross, 1973). The sources of these problems are related to adverse selection and moral hazard. Adverse selection, or "hidden information", refers to a situation where an agent has information that the principal does not have and this piece of information affects the principal's decision making (Arrow, 1985). The principal has no way of knowing whether the agent has used this information in a manner that best serves the interests of the principal. With moral hazard, or "hidden action", it is difficult for the principal to observe and assess all the agent's actions (Arrow, 1985).

Ever since Jensen and Meckling (1976) highlighted the price of agency costs to corporations, numerous scholars have brought up many tools and methods to solve this problem, such as limiting contracts to protect the creditor, or reducing incentives for incorrect behavior on the part of management personnel through competitively managing the human resources market (Fama, 1980). These also have practical uses in the realm of economics and finance, because the goals of the principal and the agents clash, supervision costs, the agents' self-restraint costs, and remaining losses result (Jensen & Meckling, 1976). The self-restraint incurred by agents in order to gain the trust of management is referred to as self-restraint costs, while any further costs created by differences in self-interest on the part of the principal and agents are referred to as the remaining loss.

Motivation

Urban (2010) considers motivation as the force that impacts human behavior in terms of completing directions. Therefore, motivation includes psychological mechanisms, and individual differences are used to explain employee behavior on the job in terms of effectiveness, satisfaction, well-being, and performance (Benedetti et al., 2015; Cerasoli et al., 2014). It moves employees from boredom to interest and encourages them to act for and accomplish specific or unmet needs. For organizations, moreover, the motivation mechanism of employees is not only the managerial means to construct stability in the organizations (Viitala & Kantola, 2016) but also the drive to increase their participation and productivity (Zhang et al., 2015). Botvinick and Braver (2015) suggest that motivation plays a critical role in personal behavioral control by increasing the incentives concerning the performance of tasks. It provides the direction for organizations to reduce the uncertainty of agency employees' performance without putting more effort into monitoring them. Moreover, rewards and encouragement might present an opportunity for employees to obtain permanent jobs based on the evidence attesting to their great performance and managers' impression. Acquiring motivation shows the work value of how agency employees might engage in cooperating with social support, and what they could contribute to the organization (Barrick et al., 2013).

According to motivation theory, there are various types of motivation associated with acceptance behavior, such as achievement motivation (Hart et al., 2004), trait motivation (Hirschfeld et al., 2004), autonomous motivation (Reinholt et al., 2011) and pro-social motivation (Grant & Berry, 2011). These different types of motivation differ in their efficiency when it comes to promoting behavior (Deci & Ryan, 2000). Employees could be motivated either intrinsically or extrinsically. Porter and Lawler (1968) propose that the two types of motivation have an additive relationship, i.e. that the two types of motivations combined lead to the highest level of motivation. However, exposing the intrinsic and extrinsic motivation of employees' behavior on the job reveals different managerial strategies for managers to focus on alternative or simultaneous motivational governance to enhance job performance (Chua et al., 2009; Robins et al., 2002).

HYPOTHESES

Having introduced the literature concerning agency theory and motivation, this study presents several hypotheses. These hypotheses are focused on the agency problem and agency cost derived from the use of temporary agency work by the user company, as well as the moderating effect of intrinsic motivation on the relationship between agency problem and agency cost.

The Relationship between Extrinsic Motivation and Agency Cost

The principal has to solve agency problems so that their benefits are not damaged. According to Rutherford et al. (2007) consideration of asymmetrical information, the principal usually has to face extremely high costs to obtain complete knowledge either by directly monitoring the agent's action or observing its unique information. However, the principal still cannot easily or precisely control the agent. From the principal's perspective, the principal monitors the agent through the organization's management mechanism or contract; therefore, the agent is willing to pursue the maximum benefits for the principal. This is called the monitoring cost and basically means the gathering of additional information about the organization's current and future economic possibilities and other information needed by the principals.

In other words, agency theory defines and prescribes a set of relationships in an organization, and the contract is the monitoring basis for the relationship between the organization and employees. Contracts draw on a range of indicators of employment and offer managers a way to monitor employees' efforts, such as attitudes and behaviors. Besides, it reflects an aspiration that employees should understand, identify with, and commit themselves to the objectives of the organization they work for (Emmott, 2015). Agency theory suggests that employee effort should be closely monitored. However, it is difficult to estimate employees' behavior, no matter how closely they are observed; the more intangible the effort, the harder this will be. How to drive employees' efforts is a critical issue for organizations. Motivation associated with performance is the concern. Greenberg and Baron (2000) note that motivation involves a set of processes that encourages, directs, and maintains personal behavior in seeking to realize the goals of organizations. It is motivating only to the extent that an individual believes that attaining the incentive is instrumental in obtaining other things of value (Vroom, 1964; Cerasoli et al., 2014), including pay, job security, and welfare. In particular, economic interests are the key driver of employees' behavior. Vansteenkiste et al. (2006) highlight that extrinsic rewards are often used as instruments of social control, and so can leave people feeling like pawns subject to the power of rewards and thus defeat their basic needs for autonomy. Organizations offer job promotions, monetary benefits, and welfare to employees directly producing the organization's benefits and performance to ensure that no shirking occurs.

Hypothesis 1: Extrinsic motivation is negatively related to an organization's agency cost.

Extrinsic Motivation, Agency Problems, and Agency Cost

An employment relationship exists when the people hired are working for the organization and agency problems are subsequently incurred (Ross, 1973). The problems arise from the effects of the self-interest assumption of agency theory. Assuming the organization (principal) and employees (agents) are self-interested utility maximizers, the problem arises from goal conflict, information asymmetry, and risk sharing between the organization and employees. The principal has to solve agency problems so that the principal's benefits are not damaged. According to Rutherford et al. (2007) consideration of asymmetrical information, the principal usually spends huge sums to obtain complete knowledge either by directly monitoring the agent's action or observing its unique information. However, the principal still cannot easily or precisely control the agent.

From the organization's perspective, the organization monitors the employees through the organization's management mechanism or contract; therefore, the employees are willing to pursue the maximum benefits for the organization. This is called the monitoring cost; it basically means gathering additional information about the organization's current and future economic possibilities and other information needed for the principals. The monitoring cost includes the expenses resulting from monitoring the employees' activities (such as bonuses, reviews, and budgets), compensation, and other costs to control employee behavior. Unavoidable agency costs are still incurred even after the organization has implemented the management mechanism is called the residual cost. It includes the adjustment cost for replenished manpower (for maternity leave), from communication or coordination.

Within the context of this research, the agency problems are focused on the organization and employees. Extrinsic motivation refers to economic interests and is the critical driver of employee behavior (Evans & Tourish, 2016). Extrinsic motivation may need to directly produce an organization's benefits and good performance by giving job promotions, monetary benefits and welfare to employees.

Based on the assumption of self-interest, the employees would be striking by releasing the benefit of their organization. The agency problems characterize the corporate governance choices of organizations and the resulting behavior of employees (Jensen & Meckling, 1976). When an organization does not have perfect information about employees' behavior, self-interested employees may conceal selfish actions, and the organization will bear the cost.

Based on agency theory, in today's world, organizations can easily diversify their responsibility through shifting employment from an internal to an external mode. Employees, on the other hand, are viewed as risk-averse and opportunistic. Employees' risk aversion stems from the fact that they are faced with the risk of losing their jobs. Aligning interests between organizations and employees seemingly makes sense, but the outcome-based incentives and multi-employment could entail unfairness and lead to increased risk levels.

According to the description of motivation, overall, we hypothesize the following:

Hypothesis 2: An organization's agency problems mediate the relationship between extrinsic motivation and the organization's agency cost.

Moderated Mediation Model

We go one step further to suggest that while extrinsic motivation provides the fundamental basis upon which organization handle agency cost through intrinsic motivation, organizations must also have access to intrinsic motivation to facilitate agency problems. Prior research highlights the motivation for improved performance or organization outcome. In particular, intrinsic motivation which results in high-quality learning, creativity and performance has emerged as an important phenomenon for organizations (Ryan & Stiller, 1991). Extrinsic motivation focuses on rewards associated with work, and reflects the gaining of short-term targets (Kohn, 1993). Consequently, possessing a certain level of motivation is more likely to drive the social behavior of employees better than competitors can, which is fundamental to obtaining sustainable performance by employees. Extrinsic motivation focuses on rewards associated with work (e.g. wages), and reflects reaching short-term targets (Kohn, 1993). We conceptualize intrinsic motivation as the substantial driver (e.g. satisfaction, freedom, a sense of accomplishment) that organizations have to invest in long-term performance.

If employees are intrinsically motivated, find their job interesting and experience enjoyment, they will work towards achieving the firms' objectives or goals (Cerasoli et al., 2014). Intrinsic motivation provides the willingness and interest for employees to work enjoyably without considering engaging in opportunistic behavior to decrease the firm's benefits or performance, and thereby reduce the costs of monitoring what workers do, because they would automatically take their job seriously and perform well. Accordingly, they may be hindered from achieving superior performance when restricted from engaging in agency problems and lack the intrinsic motivation to reduce agency costs. Therefore we offer the following:

Hypothesis 3: The indirect effect of extrinsic motivation on an organization's cost through its agency problems is moderated by intrinsic motivation.

RESULTS

A hierarchical regression approach was adopted to test the hypotheses (Cohen et al., 2003). The results of the hierarchical regression analysis in testing the hypothesized relationships among extrinsic motivation, organization agency problems, organization agency cost, and intrinsic motivation.

The first hypothesis stated that extrinsic motivation is positively associated with organization agency cost. To test the hypothesis, we conducted a hierarchical regression analysis. The control variables (industry type, working department and type of employment) were entered in the first step, and the extrinsic motivation entered in the second step. The analysis was conducted using standardized variables (Aiken & West, 1991) to facilitate interpretation and to minimize problems of multicollinearity. Hypothesis 1 was supported.

Hypothesis 2 proposes that organization agency problems (including risk sharing, information asymmetry, and goal conflict) mediate the relationship between extrinsic motivation and organization agency cost. We adopted the procedure recommended by Baron and Kenny (1986) to test this mediation hypothesis by fulfilling the following four conditions: (1) the independent variable is significantly related to the mediator; (2) the independent variable is significantly related to the dependent variable; (3) the mediator is significantly related to the dependent variable; (4) when the mediator is present, the relationship between the independent and the dependent variable becomes nonsignificant. Hypothesis 2 is supported.

Hypothesis 3 proposes that the indirect effect of extrinsic motivation on organization agency cost through organization agency problems is moderated by intrinsic motivation. According to Muller et al. (2005), the moderated mediation hypothesis is supported if three conditions are met: (1) the effect of the independent variable on the dependent variable is significant; (2) the effect of the interaction between the independent variable and the moderating variable on the mediating variable is significant, and (3) the effect of the mediating variable on the dependent variable is significant. Consequently, Hypothesis 3 is supported.

After the testing hypotheses, Hypotheses 1, 2 and 3 were supported.

DISCUSSION

The aim of this research was to test a moderated mediation model in which the indirect effect of extrinsic motivation on organization agency cost through organization agency problems was moderated by intrinsic motivation. Determining how to establish and maintain an organization's competitive advantages is an important goal given the rapid changes in international competitive trends and technological pace in the twenty-first century. Various types of employment offer maximization of organizational flexibility and maintenance or enhancement of employee performance. While the traditional and internal requirements of employment change with external demand in the design of human resources, the employee–organization relationship (EOR) is critical. When the organization requests employees to execute a certain action and provide a certain performance, an employee (agent)–organization (principal) relationship results (Jensen & Meckling, 1976). Unlike traditional agent theory, which focuses on contracts while ignoring social norms, the social perspective of agency theory manages EOR through a wide range of assumptions. All details are not contractually determined because contracts may be incomplete (Falk et al., 1999). Wiseman et al. (2012) proposed to improve the principal–agent relationships embedded in a particular social context where institutional conventions and social norms may encourage different interests; this approach should enhance the explanatory power of agency theory. Social context includes the institutional environment dimension, cognitive framework, social networks, and power relations (Wiseman et al., 2012). For example, some cultures encourage personal responsibility and concern for others over self-interest; thus, employees may pursue the honor of the organization rather than individual interest in a collectivist environment. Unavoidably, organizations that use internal and external labor face the external social context and the agent (employee)–principal (organization) contract may limit agent opportunism or influence control mechanisms. Thus, we look at EOR through the social perspective of agency theory. The word "motivation" is defined as the force that induces employees to act in a certain way to realize the fulfillment of the desired organizational performance. Organizations can motivate employees to make them work for the organization's benefit. Importantly, the motivation associated with the desired outcomes in an organization influences the performance of employees (Vallerand, 1997). We adopt motivation as the social context that increases challenges and responsibility, opportunities for advancement, and personal growth. In the following sections, we discuss the results obtained according to the formulated hypotheses.

We found that extrinsic motivation and organization agency cost are negatively related in the EOR. This result supported Hypothesis 1, which stated that extrinsic motivation can motivate employees. A possible way to enhance extrinsic motivation is through the pay received by employees (Judge et al., 2010). When an organization offers satisfactory pay to its employees, the employees may increase their performance and commitment to the organization through the operation of psychological processes (Kuvaas, 2006). Extrinsic motivation such as money, rewards, and any "pay-offs" signal to employees that their value to the organization is recognized, thereby increasing employees' desire to commit their efforts to the organization (Porter et al., 2016).

In summary, the motivation that exists at the global, contextual, and situational levels is the force that induces employees to act in a way that will fulfill the desired organizational performance. Research on motivation affects how the organization drives employee behavior regarding organizational performance. For example, Bauer et al. (2016) examined how motivation types (intrinsic motivation, motivation to learn, motivation to transfer, expectancy motivation, and task value) are related to motivation-training outcomes. Researchers suggest that employees affect organizational performance through two types of motivation (extrinsic motivation and intrinsic motivation). In most employment relationships, extrinsic motivation occurs when the organization agrees to pay the employees well, thus inspiring employees to obey the orders of the organization. Intrinsic motivation is similar to pay level, while communication with an employee regarding their potential for promotion may give the employee a sense of value, thereby resulting in greater motivation to enhance organization performance (Kondratuk et al., 2004). Accordingly, given our findings, we advise organizations to use extrinsic and intrinsic motivation according to the social perspective of the agency theory of EOR. First, the majority of current agency theory literature focuses predominantly on reducing agency problems and agency costs between shareholders and managers via an economic context, but a discussion on promoting organizational performance between the organization and employees is lacking (Shi et al., 2017; Wisemen et al., 2012). Also, some performance evaluations ignore the social context that may account for the results.

Extrinsic and intrinsic motivation may reduce the agency problems that arise from conflicts between the organization and employees. Extrinsic motivation is argued to vary considerably in its relative autonomy and thus may or may not reflect external control, and intrinsic motivation reflects the employees' propensity to learn and assimilate. In this approach, we make progress toward managing EOR with regard to the theoretical gaps and practical guidance.

REFERENCES

Arrow, K. J. (1985), The economics of agency. In: Pratt & Zeckhauser, 1985, 37-51, Principals and agents: The structure of business. Harvard Business School Press.

Barrick, M. R., Mount, M. K., & Li, N. (2013), The theory of purposeful work behavior: The role of personality, higher-order goals, and job characteristics. Academy of Management Review, 38 (1): 132-153.

Bauer, K. N., Orvis, K. A., Ely, K., & Surface, E. A. (2016), Re-examination of motivation in learning contexts: Meta-analytically investigating the role type of motivation plays in the prediction of key training outcomes. Journal of Business and Psychology, 31 (1): 33-50.

Benedetti, A. A., Diefendorff, J. M., Gabriel, A. S., & Chandler, M. M. (2015), The effects of intrinsic and extrinsic sources of motivation on well-being depend on time of day: The moderating effects of workday accumulation. Journal of Vocational Behavior, 88: 38-46.

Botvinick, M., & Braver, T. (2015), Motivation and cognitive control: From behavior to neural mechanism. Annual Review of Psychology, 66: 83-113.

Chua, J. H., Chrisman, J. J., & Bergiel, E. B. (2009), An agency theoretic analysis of the professionalized family organization. Entrepreneurship Theory and Practice, 22 (2): 355-372.

Cohen, J., & Cohen, P. (1983), Applied multiple regression/correlation analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Deci, E. L., & Ryan, R. M. (2000), The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11: 227-268.

Eisenhardt, K. M. (1989), Agency theory: An assessment and review. Academy of Management Review, 14 (1), 57-74.

Evans, S., & Tourish, D. (2016), Agency theory and performance appraisal: How bad theory damages learning and contributes to bad management practice. Management Learning, October: 1-21.

Falk, A., Gachter, S., Kovacs. (1999), Intrinsic motivation and extrinsic incentives in a repeated game with incomplete contracts. Journal of Economic Psychology, 20: 251-284.

Fama, E. F. (1980), Agency problems and the theory of the firm. Journal of Political Economy, 88 (2): 288-307.

Filatotchev, I., & Wright, M. (2011), Agency perspectives on corporate governance of multinational Enterprises. Journal of Management Studies, 48 (2): 471-486.

Gerhart, B., & Fang, M. (2014), Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects. Human Resource Management Review, 24 (1): 41-52.

Grant, A. M., & Berry, J. W. (2011), The necessity of others is the mother of invention: Intrinsic and prosocial motivations, perspective taking, and creativity. Academy of Management Journal, 54 (1): 73-96.

Greenberg, J., & Baron, R. A. (2000), Behavior in organizations, Seventh Edition. NJ: Prentice Hall.

Huselid, M. A. 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of Management Journal, 38 (3): 635-672.

Jensen, M. C., & Meckling, W. H. (1976), Theory of the organization: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, 3 (4): 305-360.
Judge, T. A., Piccolo, R. F., Rodsakoff, N. P., Shaw, J. C., & Rich, B. L. (2010), The relationship between pay and job satisfaction: A meta-analysis of the literature. Journal of Vocational Behavior, 77 (2): 157-167.
Kohn, A. (1993), Why incentive plans cannot work. Harvard Business Review.
Kondratuk, T. B., Hausdorf, P. A., Korabik, K., & Rosin, H. M (2004), Linking career mobility with corporate loyalty: How does job change relate to organizational commitment? Journal of Vocational Behavior, 65 (2): 332-349.
Kuvaas, B., & Dysvik, A. (2010), Does best practice HRM only work for intrinsically motivated employees? The International Journal of Human Resource Management, 21 (13): 2339-2357.
Porter, L., & Lawler, E. (1968), Management attitudes and performance. Homewood, Ill.: Irwin-Dorsey.
Robbins, T. L., Crino, M. D., & Fredendall, L. D. (2002), An integrative model of the empowerment process. Human Resource Management Review, 12 (3): 419-443.
Ross, S. A. (1973), The economic theory of agency: The principal's problem. American Economic Review, 63 (2): 134-139.
Rutherford, M. A., Buchholtz, A. K., & Brown, J. A. (2007), Examining the relationships between monitoring and incentives in corporate governance. Journal of Management Studies, 44 (3): 414-430.
Shi, W., Connelly, B. L., & Hoskisson, R. E. (2017), External corporate governance and financial fraud: Cognitive evaluation theory insights on agency theory prescriptions. Strategic Management Journal, 38: 1268-1286.
Tsui, A. S., & Wu, J. B. (2005), The new employment relationship versus the mutual investment approach: Implications for human resource management. Human Resource Management, 44: 115–121.
Tsui, A. S., Pearce, J. L., Porter, L. W., & Hite, J. P. (1995), Choice of employee-organization relationship: Influence of external and internal organizational factors. Personnel & Human Resource Management, 13: 117-151.
Tsui, A. S., Pearce, J. L., Porter, L. W., & Tripoli, A. M. (1997), Alternative approaches to the employee-organization relationship: Does investment in employees pay off? Academy of Management Journal, 40: 1089–1121.
Urban, B. (2010), Cognitions and motivations for new venture creation decisions: Linking expert scripts to self-efficacy, a South African study. International Journal of Human Resource Management, 21 (9): 1512-1530.
Viitala, R., & Kantola, J. (2016), Temporary agency workers shake a work community: a social capital perspective. Employee Relations, 38 (2):147-162.
Whitener, E. M., Brodt, S. E., Korsgaard, M. A., & Werner, J. M. (1998) Managers as initiators of trust: An exchange relationship framework for understanding managerial trustworthy behavior. Academy of Management Review, 23: 513-530.
Wiseman, R. M., Cuevas-Rodriguez, G., & Gomez-Mejia, L. R. (2012), Toward a social theory of agency. Journal of Management Studies, 49 (1): 202-222.
Zhang, A. Y., Tsui, A. S., Song, L. J., Li, C., & Jia, L. (2008), How do I trust thee? The employee-organization relationship, supervisory support, and middle manager trust in the organization. Human Resource Management, 47 (1): 111–132.

DECISION SCIENCES INSTITUTE

Exploring Pollution Levels During COVID Pandemic

Dina Gisel
DePaul University
dallen22@depaul.edu

Mark Betti
DePaul University
mbetti@depaul.edu

Tess Gorman
DePaul University
tgorman7@depaul.edu

Seif Abuhashish
DePaul University
sabuhash@depaul.edu

Ilyas Ustun
DePaul University
iustun@depaul.edu

ABSTRACT

This study examines the potential link between pollution and COVID-19 transmission and mortality in the United States. While various pollutants are associated with respiratory diseases, this analysis reveals that pollution, although contributing to health issues over time, has limited direct impact on COVID-19 deaths. Our findings indicate that pollution levels, when lagged two weeks behind in a Time Series Analysis, show some correlation with COVID-19 deaths. However, a linear regression model using pollutant levels as predictors has limited predictive power, accounting for less than 4% of the data's variation. It is clear that other factors play a more substantial role in COVID-19 outcomes. Expanding the model and considering additional variables is essential for a more comprehensive understanding.

KEYWORDS: Air Pollution, COVID-19, Environmental Analysis, Air Quality Analysis, and Explanatory Visualization

INTRODUCTION

Climate change and global warming are important issues that negatively affect life across our planet. The World Health Organization estimates that between the years 2030 and 2050, climate change will increase the number of deaths per year (World Health Organization 2021). In addition to the loss of human life, the organization estimates that by 2030, the cost of climate change per year will be between $2 and $4 million (World Health Organization 2021). The impact of climate change is felt around the world, with more extreme consequences for developing countries (World Health Organization 2021). Global warming can have detrimental impacts on our planet through changing ecosystems and increasing the likelihood and size of extreme weather events, which are increasingly costly and harmful to human and animal life (DCCEEW).

The emission of pollutants has detrimental effects on air quality and produces climate change, which negatively impacts ecosystems around the world. Carbon monoxide emissions impact greenhouse gas levels in the earth's atmosphere, which contribute to climate change and global warming (DCCEEW). Increased concentrations of sulfur dioxide harm plant matter and contribute to acid rain, which can have negative impacts on the human respiratory system. Ozone is the third most important greenhouse gas and exists in both the stratosphere and the troposphere (Environmental Protection Agency 2023a). In the stratosphere, ozone is helpful in reducing the level of ultraviolet radiation that reaches Earth (Environmental Protection Agency 2021). However, in the troposphere, ozone has negative effects on human health. Tropospheric ozone abundance is primarily influenced by gas emissions of methane, carbon monoxide, and nitrogen oxides (Environmental Protection Agency 2023a). The Clean Air Act established National Ambient Air Quality Standards (NAAQS) for six major air pollutants that have proved to be detrimental to public health and the environment (Environmental Protection Agency 2023b).

Analyzing the levels of pollutants across the US is important to understanding where the effects of climate change might be felt the hardest across the US and where air quality-caused health problems might be the highest. With the COVID-19 virus being a respiratory illness and the pandemic halting travel across the US, analyzing variations in pollution as well as the effect of pollution on the pandemic outcomes is vital for public health policy and pollution regulations to slow climate change.

LITERATURE REVIEW

Effects of Pollution on American Health

The Clean Air Act (CAA 1970) mandated health-based national ambient air quality standards (NAAQS) set at a level "requisite to protect public health with an adequate margin of safety." Health impact assessments (HIAs) are used to make informed and systematic decisions about regulatory policy based on estimates of positive and negative health impacts from proposed air pollution standards (Briggs et al. 2009). It has even been demonstrated that long-term exposure to air pollution is associated with an increased prevalence of respiratory diseases and deaths (Brauer M 2010). As of now, fine particulate matter with size <2.5 μm (PM2.5) is considered to be one of the major health risk factors in the environment (Lelieveld et al. 2015). Not only that, NO2, a traffic-related air pollutant, is also linked with adverse health outcomes, although it is often not quantified in pollution-attributable disease burden studies, potentially because coarsely resolved concentration estimates are often unable to capture highly spatially variable patterns in NO2 (Anenberg et al. 2017). Some studies have been able to resolve this dilemma through mobile monitoring and modeling, complemented by satellite remote sensing. These measurements have been used to create street-level annual average concentrations of NO2 and BC, a land-use regression (LUR) model (Messier et al. 2018)

COVID's Impact in America

At its peak, more than 113 million people worldwide had been diagnosed with COVID-19, resulting in more than 2.5 million deaths (World Health Organization, 2021). In the United States, factors such as age, race/ethnicity, and other sociodemographic characteristics appear to increase risk for COVID-19 infection, severity, and associated death (Brandt et al., 2020). For example, compared to non-Hispanic whites, cumulative COVID-19 hospitalization rates for Black and Latino populations are approximately 4.7 and 4.6 times higher in the U.S., respectively (Centers for Disease Control and Prevention, 2020). This also can be factored in by Air Pollution. Exposure to air pollution is considered as the major environmental cause of several diseases and premature death around the globe (Global Burden of Disease 2017). Some reviews highlighted the links between air pollution and COVID-19 however, up to now, a limited number of data-dependent studies have been conducted to investigate the association between air pollution and COVID-19 infection and mortality (Huraimel et al. 2020, Comunian et al. 2020, Copat et al. 2020, Domingo et al. 2020). The available aforementioned studies have demonstrated the effects of short-term (within 2 months of exposure) and long-term exposure (more than 2 months of exposure) to air pollution on COVID-19 infections and mortality.

Health Benefits from Large Scale Ozone Reduction

The Clean Air Act of 1970 mandated air quality standards in an effort to protect public health given that tropospheric ozone is related to decreased pulmonary function, asthma, and increased hospital and emergency department visits. Data from the EPA that looked at pollution levels between the years 2005 and 2007 was used to estimate the potential changes in ozone and their effects on mortality and morbidity. The study that worked on analyzing this data looked at various levels and what their resulting impact would be. Reducing the regulated levels to 60 ppb increased prevented mortalities by 3.5 times compared to the 75 ppb regulatory level (Berman et. Al 2012). Areas in the US with high levels of ozone, such as Southern California, Industrial Midwest, portions of the south, St.Louis, and Atlanta, benefited more from adjustments to regulated standards. Prior to 2015, the standard was set to 75 ppb but after the work done to prove that lower ppb standards would result in ozone reduction, the regulated levels were adjusted to 70 ppb moving the US to cleaner air with lower ozone levels.

Effect of Pollution Exposure with COVID

In previous studies, it has been identified that air pollution is associated with an increased prevalence of respiratory disease and deaths. Long-term exposure, defined as more than 2 months, to air pollution is associated with an increased prevalence of respiratory disease and death. The transmission and infection of COVID-19 behaved in a similar manner to other respiratory diseases. While pollution plays an impact, there are various confounding effects such as gender, age, smoking, and population density that all impact higher morbidity and mortality in relation to COVID and therefore caution should be taken when translating pollution values to vulnerability (Ali N., Islam F. 2020). Correlation between infection and mortality compared with various population densities results in positive, neutral, and negative correlations. Where positive and neutral correlation can be assumed, higher density results in higher transmission and therefore higher deaths, the negative correlations could be explained by people moving to country-sides during the COVID 19 lockdowns. This made studying pollution's true effect on COVID-19 patients more difficult given a single person's actual exposure to pollution during their COVID diagnosis. Zhou et al. (2021) investigated the impact of fine particulate matter (PM2.5) exposure during the 2020 wildfires in the United States on COVID-19 cases and deaths. They found that wildfires increased the effect of short-term PM2.5 exposure on COVID-19 outcomes, although with variability across counties. Lipsitt et al. (2021) focused on Los Angeles and examined the association between air pollution and COVID-19 incidence and mortality. Their findings suggested that chronic exposure to nitrogen dioxide had significant effects on COVID-19 disease cases and mortality, particularly in neighborhoods with higher levels of air pollution and larger proportions of Latinx and Black populations. Overall, many authors have undertaken the research efforts to understand the effects of air pollution on COVID outcomes.

DATA

To perform this analysis, we merged two different datasets. The first dataset was sourced from the US EPA and contains daily, county-level pollutant level measurements for four major gas pollutants that impact air quality: carbon monoxide, nitrogen dioxide, ground level ozone, and sulfur dioxide. It also contains daily measurements for particulate matter 2.5 and particulate matter non-2.5. The summary statistics of these data can be seen in Figure 1. The pollutant measured values include the mean for a given day and the hour where the levels reached their max. The second dataset from the New York Times contains daily, county-level data for COVID deaths and case counts. The data range we used was from 1/21/2020 through 11/1/2022.

Figure 1: Statistical Summary

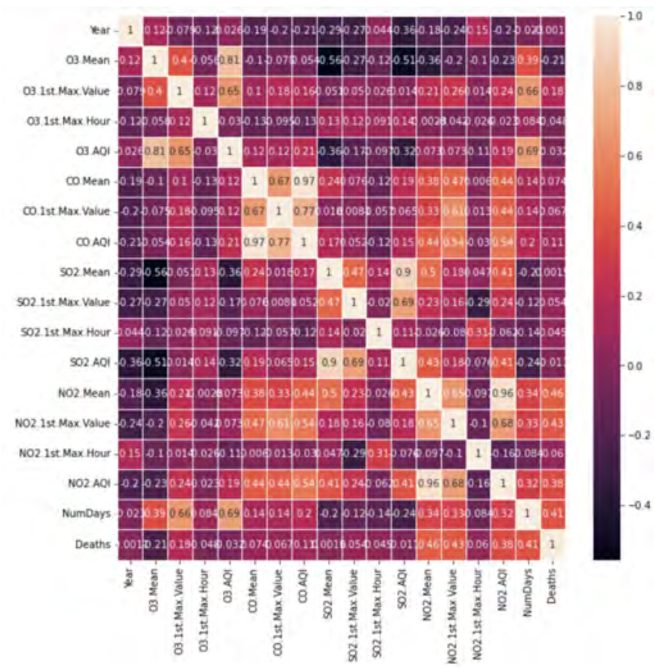
Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean	Std. Deviation
CO_Arithmetic.Mean	134469	-4.00000E-1	1.58024E+1	3.44666E+4	2.563160E-1	1.708957E-1
CO_1st.Max.Value	134469	-.40	29.00	68114.04	.51	.43
NO2_Arithmetic.Mean	134469	-2.87	55.79	1144371.14	8.51	5.65
NO2_1st.Max.Value	134469	-1.50	101.60	2728522.40	20.29	11.81
SO2_Arithmetic.Mean	134469	-1.4463E+0	3.26688E+1	6.04680E+4	4.496795E-1	6.131989E-1
SO2_1st.Max.Value	134469	-1.00	564.50	211249.80	1.57	3.83
Ozone_Arithmetic.Mean	134469	-3.3000E-3	7.7600E-002	4.0968E+3	3.046659E-2	9.918710E-3
Ozone_1st.Max.Value	134469	-.003	.183	5505.956	.041	.013
PM25_Arithmetic.Mean	134469	-3.112	334.000	1113432.75	8.280	6.561
PM25_1st.Max.Value	134469	-1.00	999.90	2537987.30	18.87	19.92
PM25Non_Arithmetic.Mean	134469	-3.973	426.865	1071058.08	7.965	4.878
PM25Non_1st.Max.Value	134469	.00	985.00	2487709.80	18.50	12.83
Covid.Deaths	134469	.00	1462.00	343620.00	2.56	12.85
Covid.Cases	134469	.00	110441.00	34323492.00	255.25	1243.98
Valid N (listwise)	134469					

We created a correlation heatmap to observe the relationship among the pollutants. We found that the number of deaths is moderately positively correlated with NO2 Mean (0.46), NO2 1st Max Value (0.43), and NO2 AQI (0.38). On the other hand, O3 Mean is moderately negatively correlated with SO2 mean (-0.56) and SO2 AQI (-0.51). We also discovered that CO Mean is moderately positively correlated with NO2 AQI (0.44) and CO 1st Max Value is correlated with

NO2 1st Max Value (0.61). These correlations are similar to those found between SO2 and NO2 means and AQIs which are shown in Figure 2.

Figure 2: Correlation between variables



Data Cleaning

Data cleaning involved removing rows and counties where all data wasn't present. This was particularly important given that a times series analysis was part of the study. To avoid breaks in the data, full records needed to be available. Some instances of missing data were imputed given that a value of NA meant there was no data available yet. This was prevalent with COVID case and death reporting in early months given COVID exposure hadn't reached all counties yet. The COVID data was also on an aggregate basis. In order to evaluate, we adjusted the data to be incremental to analyze the increase of case and death counts as the days went on. Rows that still held an NA attribute were replaced with the columns average to avoid breaks when conducting analysis. These changes and adjustments resulted in a 4.38% reduction in total data which fell into the acceptable range in regard to the remaining integrity of the data.

Data Transformation

The data transformation processes in our study involved several steps. First, we converted the original aggregate daily data to incremental data, capturing the changes and variations for each specific day. This allowed us to perform our time series analyses. Additionally, we performed data aggregation at the county level for certain statistical tests, consolidating the data to run the tests. Furthermore, to examine regional variations, we added a variable 'Region' to our data to use for some of our statistical tests.

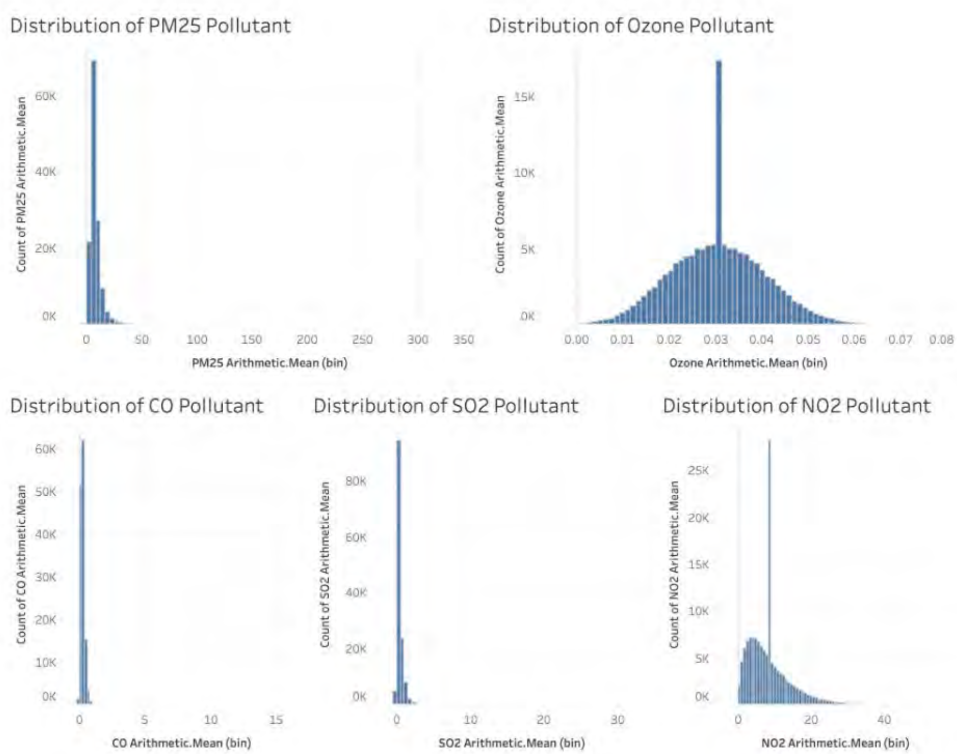
METHODOLOGY

Our study investigates the correlation between pollutant levels (CO, NO2, O3, SO2, and PM2.5) and COVID-19 cases and deaths in the United States. Initially, we conducted an exploratory analysis to compare pollutant levels over time and across different states using Tableau and SPSS. Then, we performed a statistical analysis to identify any differences in pollutant levels before and after the COVID-19 shutdown and to determine if there was any association between these levels and the number of COVID-19 cases and deaths. Finally, we conducted a regression and time series analysis to determine if any pollutant values predicted COVID-19 cases and deaths in the United States.

Exploratory Analysis

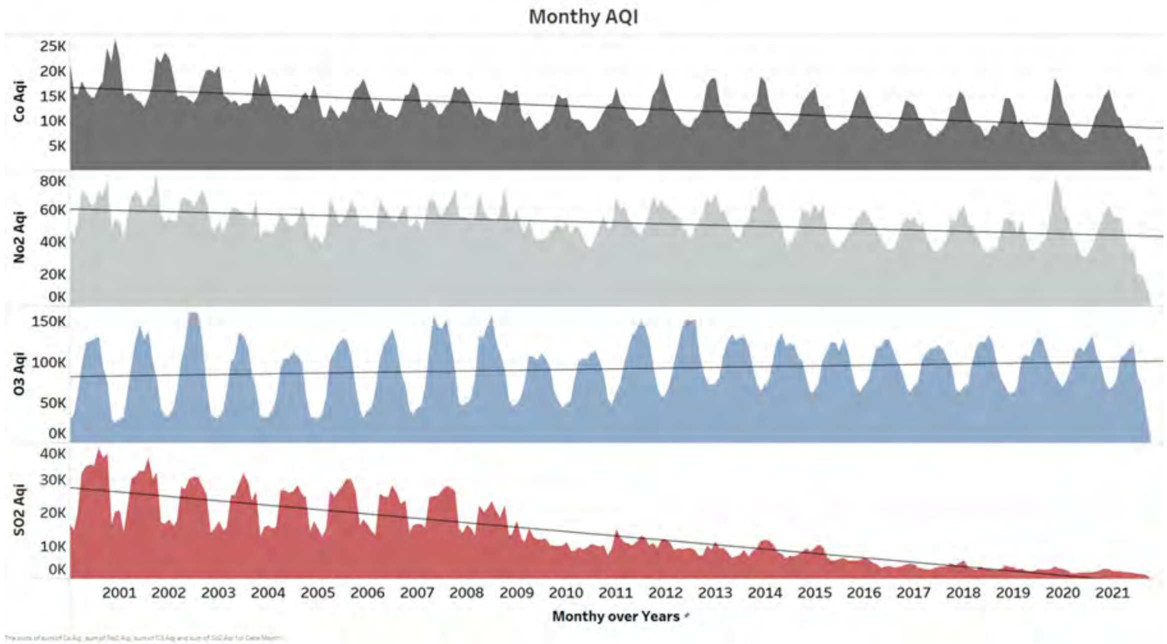
We started our analysis by examining the distribution of different pollutants using histograms as shown in Figure 3. Through this, we discovered that all pollutants, except for O_3, were skewed to the right and had noticeable outliers.

Figure 3: Distribution of pollutants



We then explored the Air Quality Index (AQI) for each of the pollutions as shown in Figure 4. Initial analysis shows the CO has transitioned to singular peaks instead of double peaks after 2010. NO2 was sporadic prior to 2010 and formed more cyclical patterns onwards, peaking between the months of November to January. The biggest observation was of O3 AQI, we noticed that the bottom of the peaks (named throughs) showed an increase over the past 20 years, with an increase of 167.12% since 2000. If we break this down further, it grew over 90.27% between 2000 to 2010 and then another 40.39% from 2011 to 2020. This implies that during the through cycle, more pollution is being produced than at its peak. Though there is one positive outlook and that is the complete reduction of SO2 AQI. SO2 has seen a steady decline after 2009. This implies that the production of SO2 has been reduced either through policy making or through good faith.

Figure 4: Monthly Air Quality Index (AQI) for each of the pollutions



Continuing on, we explored how the AQI for each of the pollutions affected each of the states within the US. From Figure 5, we noticed that California is consistently ranked the highest in AQI pollutants for CO, NO2, O3 and 2nd highest for SO2 with Pennsylvania ranking 1st. Initial thoughts regarding this high ranking for California could be tied to its geographical location, population size, and/or possible industry within the state. As for Pennsylvania, this could potentially be associated with coal processing plants for electricity generation, though there is evidence that they have been reducing their coal emission in the past 5 years (DeVilbiss & Ray 2017).

In addition, our recent analysis shows how the pollution level in a county relates to the number of COVID-19 deaths in that same county. The study uses a color-coded map where darker points represent higher pollution levels, and larger points indicate a greater number of COVID-19 deaths as shown in Figure 6. It seems that the concentration of pollutants in an area is consistently linked to the number of deaths. However, SO2 has consistently been found to have the lowest concentration levels across the country and does not appear to have an impact on the death rate in specific regions.

Figure 5: AQI by state

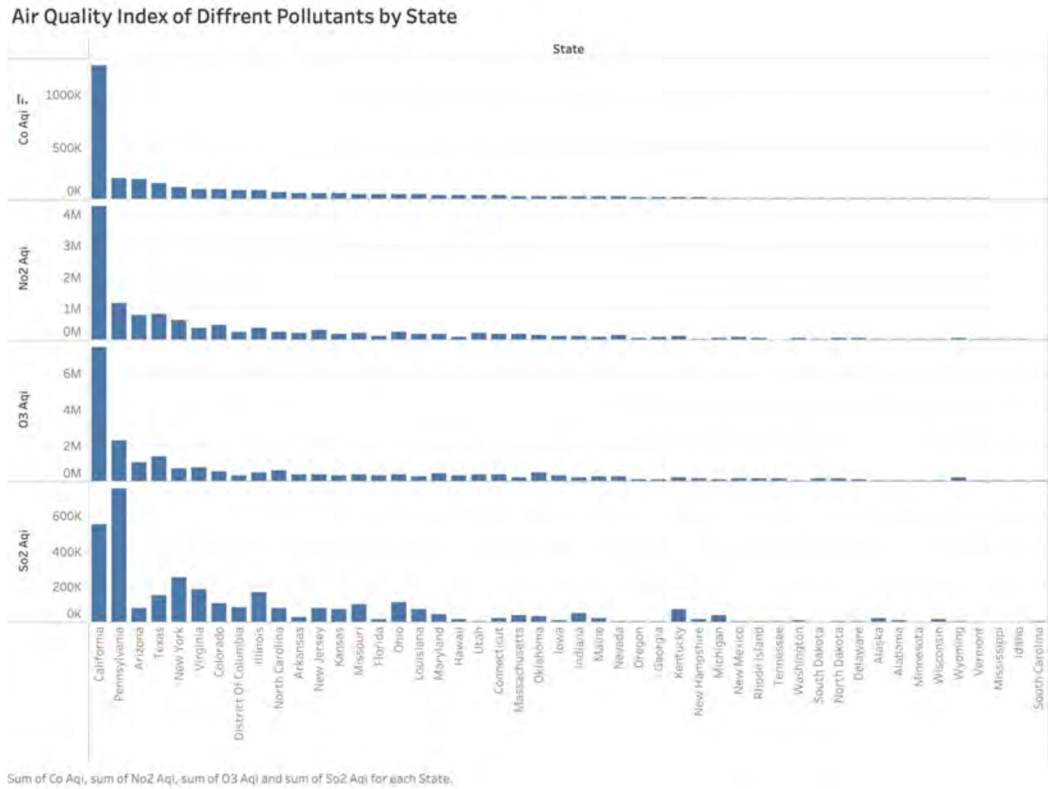
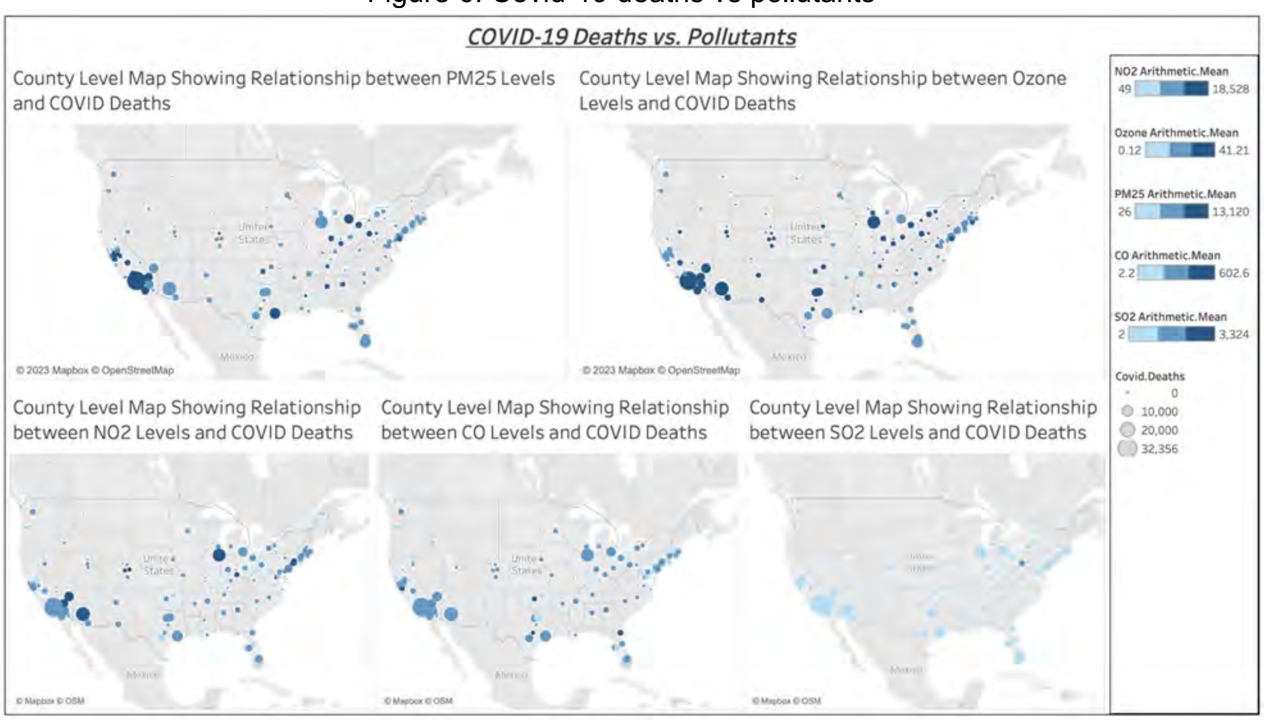


Figure 6: Covid-19 deaths vs pollutants



Statistical Tests
Our methodology incorporated a series of statistical tests designed to answer several research questions. To investigate whether there was a difference in pollutant levels pre and post shutdown, we partitioned each pollutant variable into two categories: pre and post shutdown averages. A Shapiro-Wilks test for normality was run on each set, followed by a Sign Rank paired T test, given there was at least one non-normal variable in each pair.

To examine the relationship between COVID-19 cases and ozone levels, we first tested the variables for normality. Given that neither variable exhibited a normal distribution, a Spearman Rank correlation test was used to determine the significance of the relationship.

A similar approach was undertaken to understand the relationship between COVID-19 deaths and ozone levels, as well as PM 2.5 levels. We applied the Shapiro-Wilks test for normality on the variables. Upon finding the distributions non-normal, Spearman Rank correlation tests were employed to test the significance of the correlations.

Finally, to explore regional differences in pollutant levels, we examined the normality of mean pollutant variables by region. Depending on the normality, we used either a Welch test for normal distributions or a Kruskal-Wallis non-parametric ANOVA for non-normal variables. These analyses enabled an understanding of the relationships between pollutant levels and COVID-19 deaths, as well as pre and post shutdown pollutant changes and regional variations in pollutant levels.

Regression Analysis

We created a multiple linear regression model to investigate whether pollutant levels could predict COVID-19 cases and deaths in the US. Two separate models were built for COVID-19 cases and deaths, with pollutant levels (CO, NO2, O3, SO2, and PM2.5) as the independent variables. These variables were measured in parts per billion for NO2 and SO2 and parts per million for CO, O3, and PM2.5. The pollutants were expressed as average and maximum daily values for each county and State in the USA, from January 21, 2020, to November 1, 2022. We used a stepwise regression method in SPSS to build the models. The regression analysis estimated the regression beta coefficients, p-values, and Adjusted R-squared values to evaluate the variable associations and the model's significance and accuracy.

Time Series Analysis

The data used for this study was time series in nature. Initial exploration showed seasonality in all of the time series aside from the Particulate Matter 2.5 pollutant. The goal was to determine if there was any correlation between the pollutants values on a given day and the deaths or transmission levels that followed. According to the paper summarized above titled The Effects of Air Pollution on COVID-19 Infection and Mortality (Ali, Islam,2020), long-term exposure, which is defined as longer than 2 months, has been proven to be associated with an increased prevalence of respiratory disease and death. Because of this, a lag analysis was performed which lagged covid deaths and cases and compared the lagged time series to the time series of the different pollutants.

Regarding COVID-19 Deaths, Sulfur Dioxide, Nitrogen Dioxide, and Carbon Monoxide exhibited the highest correlation values. Particularly, the pollution levels were correlated the most with COVID-19 deaths when the individual pollution series were lagged two weeks. Up to a 15-day lag was chosen to analyze a full half month as well as the two-week mark which is typically the incubation period for the disease. The lag plot should be read as lagging the pollution a certain number of days to see what the impact day 0 of pollution levels had on day 5, for example, of COVID deaths. Figures 7, 8, and 9 show the lag plots analyzed. Two conclusions can be drawn from these lag plots. The first being that while correlation was highest for these three pollutants, there isn't a specific lag that correlation values stand out and guide a deeper analysis. Correlation remains within the .2-.4 range. This indicates that while some pollutants may impact mortality more than others, pollution doesn't impact mortality significantly enough on its own. If pollution were to significantly impact mortality, we would expect to see a spike in correlation at a certain lag period which would indicate that if pollution levels are high on a given day, more deaths would occur. Since this does not occur, this indicates that there are other factors that are contributing to COVID-19 mortality.

In a study that examined the effects of air pollution on COVID-19 infection and mortality it has been identified that air pollution is associated with an increased prevalence of respiratory disease and deaths (Ali, Islam,2020). This study also concluded that there are various confounding effects such as gender, age, smoking, population density, and pollution that all impact higher morbidity and mortality in relation to COVID and therefore caution should be taken when translating pollution values to vulnerability.

Figure 7: Lag plots of deaths w.r.t. sulfur dioxide

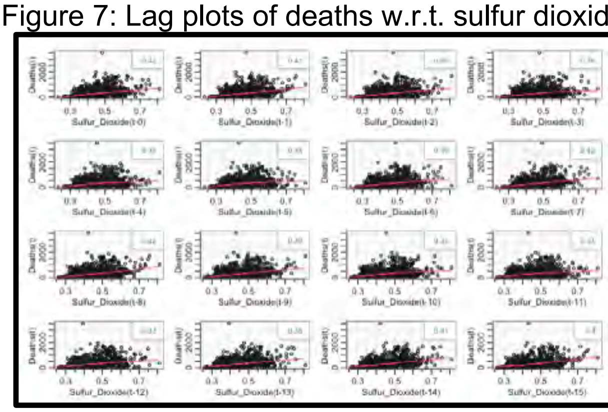


Figure 8: Lag plots of deaths w.r.t. nitrogen dioxide

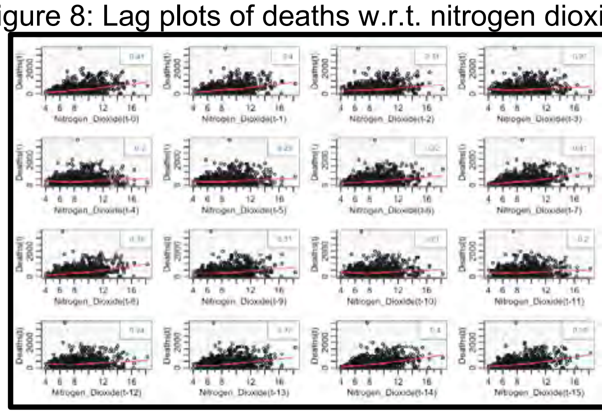
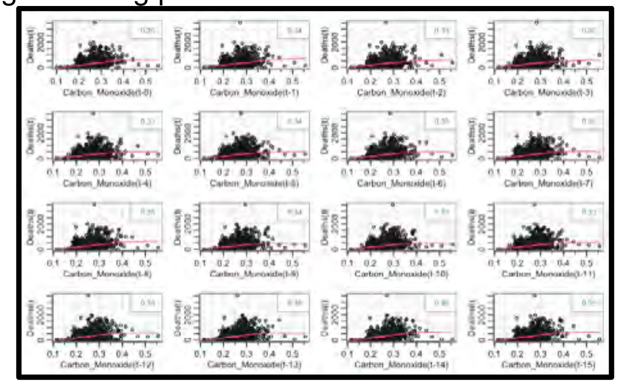


Figure 9: Lag plots of deaths w.r.t. carbon monoxide



Lag plots for COVID-19 transmission and case counts were also analyzed as shown in Figures 10, 11, and 12. When looking at transmission, Sulfur Dioxide, Nitrogen Dioxide and Carbon Monoxide similar to deaths, showed the highest correlations at the two-week mark when pollution levels were lagged. Similar to the COVID-19 death analysis, correlations are fairly uniform within the lag analysis. If pollution levels on a given day truly impacted or were predictive of cases on a given day in the future we would expect to see an increase in correlation values in the lag analysis. Given that this is not the case, we believe that while pollution may not encourage quick recovery from COVID-19, it is not the leading factor in transmission or death counts.

Figure 10: Lag plots of cases w.r.t. sulfur dioxide

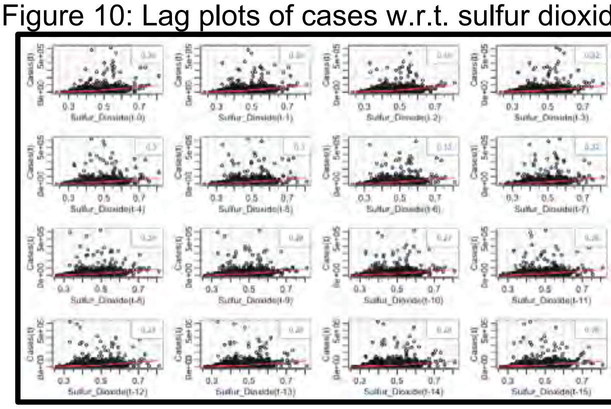


Figure 11: Lag plots of cases w.r.t. nitrogen dioxide

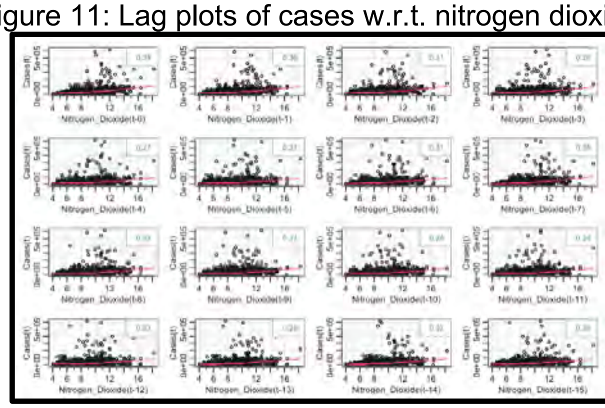
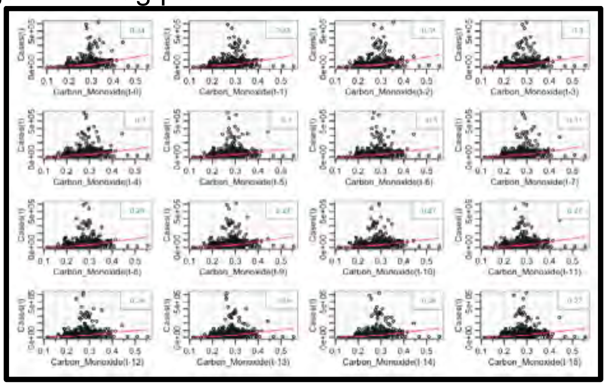


Figure 12: Lag plots of cases w.r.t. carbon monoxide



RESULTS & DISCUSSION

Statistical Tests

Our statistical tests focused on analyzing the differences in pollutant levels before and after shutdowns, relationships between COVID-19 cases, deaths, and pollutant levels, and variances in pollutant levels across regions. The Wilcoxon Signed-Rank test as shown in Figure 13 revealed statistically significant differences ($p<0.05$) in CO, NO2, Ozone, PM2.5, and Non-PM2.5 levels pre- and post-shutdown, all pollutants except for SO2.

Figure 13: Wilcoxon-signed rank test results

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	3.344125	Pr > |t|	0.0010
Sign	M	22	Pr >= |M|	0.0004
Signed Rank	S	2015	Pr >= |S|	0.0002

Test of CO Mean Levels Pre and Post Shutdown

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	13.59411	Pr > |t|	<.0001
Sign	M	56	Pr >= |M|	<.0001
Signed Rank	S	4036	Pr >= |S|	<.0001

Test of NO2 Mean Levels Pre and Post Shutdown

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	0.480294	Pr > |t|	0.6317
Sign	M	-3	Pr >= |M|	0.6660
Signed Rank	S	61	Pr >= |S|	0.8928

Test of SO2 Mean Levels Pre and Post Shutdown

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	-8.55193	Pr > |t|	<.0001
Sign	M	-37	Pr >= |M|	<.0001
Signed Rank	S	-3667.5	Pr >= |S|	<.0001

Test of Ozone Mean Levels Pre and Post Shutdown

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	-5.5774	Pr > |t|	<.0001
Sign	M	-27	Pr >= |M|	<.0001
Signed Rank	S	-2828.5	Pr >= |S|	<.0001

Test of PM2.5 Mean Levels Pre and Post Shutdown

Tests for Location: Mu0=0				
Test		Statistic	p Value	
Student's t	t	-5.09121	Pr > |t|	<.0001
Sign	M	-47	Pr >= |M|	<.0001
Signed Rank	S	-2804.5	Pr >= |S|	<.0001

Test of NonPM2.5 Mean Levels Pre and Post Shutdown

The Spearman correlation tests as depicted in Figure 14 showed no significant correlation between COVID-19 cases and deaths with ozone levels, while a significant positive correlation was found between COVID-19 deaths and PM2.5 level.

Figure 14: Spearman correlation tests results

Spearman Correlation Coefficients, N = 155 Prob > |r| under H0: Rho=0	
	Ozone_Arithmetic_Mean
Covid_Cases	-0.01637 0.8397

Correlation Test between Ozone Mean and Covid Cases

Spearman Correlation Coefficients, N = 155 Prob > |r| under H0: Rho=0	
	Ozone_Arithmetic_Mean
Covid_Deaths	0.00403 0.9603

Correlation Test between Ozone Mean and Covid Deaths

Spearman Correlation Coefficients, N = 155 Prob > |r| under H0: Rho=0	
	PM25_Arithmetic_Mean
Covid_Deaths	0.37304 <.0001

Correlation Test between PM 2.5 Mean and Covid Deaths

Finally, the Kruskal-Wallis and Welch tests shown in Figure 15 suggested statistically significant differences in pollutant levels across regions for SO2, PM2.5, and Non-PM2.5.

Figure 15: Kruskal-Wallis and Welch test results

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
4.4814	3	0.2140

Test of CO Mean Levels by Region

Welch's ANOVA for NO2_Arithmetic_Mean			
Source	DF	F Value	Pr > F
Region	3.0000	0.39	0.7632
Error	71.1441		

Test of NO2 Mean Levels by Region

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
15.2951	3	0.0016

Test of SO2 Mean Levels by Region

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
5.9464	3	0.1142

Test of Ozone Mean Levels by Region

Welch's ANOVA for PM25_Arithmetic_Mean			
Source	DF	F Value	Pr > F
Region	3.0000	4.80	0.0043
Error	68.4922		

Test of PM2.5 Mean Levels by Region

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
12.5643	3	0.0057

Test of NonPM2.5 Mean Levels by Region

The significant p-values for CO, NO2, ozone, PM2.5, and non-PM2.5 levels in the pre/post shutdown analysis suggests that shutdowns due to COVID-19 had an impact on the levels of these pollutants. However, SO2 levels remained relatively unchanged. We believe this is as expected as SO2 seems to be the pollutant that would be least impacted by a slowdown of travel. Interestingly, PM2.5 levels showed a strong positive correlation with COVID-19 deaths, indicating that higher PM2.5 concentration may be associated with higher COVID-19 fatality rates. The lack of correlation between COVID-19 metrics and ozone levels suggests that ozone might not directly influence COVID-19 outcomes. The regional analysis highlights that pollutant distributions are not the same across regions, suggesting local environmental and industrial factors impact pollutant levels, especially with regards to SO2 and Particulate Matter.

Regression Analysis

Our linear regression (LR) models showed statistical significance in predicting COVID-19 deaths and cases but had poor R^2 values. For COVID-19 deaths as our dependent variable, the ANOVA chart revealed p-values of less than .005 as shown in Figure 16, indicating the independent variables' significance in predicting our dependent variable. After optimizing our model through the linear regression stepwise method, we found that model 6 with six independent variables could predict COVID-19 deaths. However, our adjusted R-squared value of 0.032 as depicted in Figure 17, demonstrated that the model only accounted for 3.2% of the observed variation in COVID-19 deaths. This low value may have been influenced by other independent and confounding variables such as age, weight, race/ethnicity, and socioeconomic status.

Using beta coefficients, we described the estimated change in the dependent variable (COVID-19 deaths) for a one-unit increase in the corresponding independent variable (pollutant levels) while holding other variables constant. Positive beta coefficients indicate a positive relationship, meaning higher pollutant levels are associated with higher COVID-19 deaths. Negative coefficients indicate a negative relationship. Our final linear regression model as shown in Figure 18 with its respective beta coefficients is:

$$COVID.Deaths = .222 + .167(NO2\_1st.Max.Value) + 3.356(CO\_1stMax.Value) - 4.984(CO\_Arithmetic.Mean) - .136(NO2\_arithmetic.Mean) - 116.205(Ozone\_arithmetic.Mean) + 78.641(Ozone\_1^{st}.Max.Value)$$

Figure 16: LR models predicting Covid-19 deaths

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	524834.113	1	524834.113	3254.628	<.001[b]
	Residual	21682402.7	134458	161.258		
	Total	22207236.8	134459			
2	Regression	591139.030	2	295569.515	1838.509	<.001[c]
	Residual	21616097.7	134457	160.766		
	Total	22207236.8	134459			
3	Regression	646154.275	3	215384.758	1343.150	<.001[d]
	Residual	21561082.5	134456	160.358		
	Total	22207236.8	134459			
4	Regression	665195.572	4	166298.893	1037.957	<.001[e]
	Residual	21542041.2	134455	160.217		
	Total	22207236.8	134459			
5	Regression	676557.805	5	135311.561	844.989	<.001[f]
	Residual	21530679.0	134454	160.134		
	Total	22207236.8	134459			
6	Regression	709387.406	6	118231.234	739.448	<.001[g]
	Residual	21497849.4	134453	159.891		
	Total	22207236.8	134459			
7	Regression	712621.191	7	101803.027	636.793	<.001[h]
	Residual	21494615.6	134452	159.868		
	Total	22207236.8	134459			
8	Regression	715475.338	8	89434.417	559.496	<.001[i]
	Residual	21491761.4	134451	159.848		
	Total	22207236.8	134459			
9	Regression	716958.080	9	79662.009	498.391	<.001[j]
	Residual	21490278.7	134450	159.838		
	Total	22207236.8	134459			

a. Dependent Variable: Covid.Deaths

Figure 17: Model summaries of LR models predicting Covid-19 deaths

Model Summary[j]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.154[a]	.024	.024	12.699	
2	.163[b]	.027	.027	12.679	
3	.171[c]	.029	.029	12.663	
4	.173[d]	.030	.030	12.658	
5	.175[e]	.030	.030	12.654	
6	.179[f]	.032	.032	12.645	

Figure 18: Final LR model for predicting Covid-19 deaths

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-.840	.069		-12.196	<.001
	NO2_1st.Max.Value	.167	.003	.154	57.049	<.001
2	(Constant)	-1.074	.070		-15.404	<.001
	NO2_1st.Max.Value	.128	.004	.118	36.602	<.001
	CO_1st.Max.Value	2.029	.100	.065	20.308	<.001
3	(Constant)	-.456	.077		-5.905	<.001
	NO2_1st.Max.Value	.126	.004	.116	36.036	<.001
	CO_1st.Max.Value	3.966	.145	.128	27.438	<.001
	CO_Arithmetic.Mean	-6.075	.328	-.079	-18.522	<.001
4	(Constant)	-.539	.078		-6.955	<.001
	NO2_1st.Max.Value	.178	.006	.163	30.220	<.001
	CO_1st.Max.Value	3.792	.145	.122	26.085	<.001
	CO_Arithmetic.Mean	-5.092	.340	-.066	-14.978	<.001
	NO2_Arithmetic.Mean	-.132	.012	-.058	-10.902	<.001
5	(Constant)	.428	.139		3.088	.002
	NO2_1st.Max.Value	.190	.006	.175	31.366	<.001
	CO_1st.Max.Value	3.677	.146	.118	25.192	<.001
	CO_Arithmetic.Mean	-5.046	.340	-.066	-14.844	<.001
	NO2_Arithmetic.Mean	-.160	.013	-.070	-12.716	<.001
	Ozone_Arithmetic.Mean	-30.739	3.649	-.024	-8.423	<.001
6	(Constant)	.222	.139		1.592	.111
	NO2_1st.Max.Value	.167	.006	.154	26.729	<.001
	CO_1st.Max.Value	3.356	.148	.108	22.742	<.001
	CO_Arithmetic.Mean	-4.984	.340	-.065	-14.671	<.001
	NO2_Arithmetic.Mean	-.136	.013	-.060	-10.780	<.001
	Ozone_Arithmetic.Mean	-116.205	6.991	-.090	-16.622	<.001
	Ozone_1st.Max.Value	78.641	5.488	.079	14.329	<.001

When we used COVID-19 cases as our dependent variable, the ANOVA chart revealed p-values of less than .005 as shown in Figure 19 indicating the independent variables' significance in predicting our dependent variable. After optimizing our model through the linear regression stepwise method, we also found that model 6 with six independent variables could predict COVID-19 cases. Our adjusted R-squared values are shown in Figure 20. They have slightly increased to 0.038 indicating that the model only accounted for 3.8% of the observed variation in COVID-19 cases. As mentioned previously this may be due to other independent or confounding variables like age, weight, race/ethnicity, pervious health conditions and even socioeconomic status.

Our final linear regression model is shown in Figure 21. The beta coefficients for COVID-19 cases is:

$$COVID.Casess = 139.842 + 10.841(NO2\_1st.Max.Value) + 380.003(CO\_1stMax.Value) - 14108.300(Ozone\_Arithmetic.Mean) - 468.453(CO2\_Arithmetic.Mean) + 6709.412(Ozone\_1^{st}.Max.Value) - 48.353(SO2\_Arithmetic.Mean)$$

Figure 19: LR models predicting Covid-19 cases

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5424213351	1	5424213351	3598.722	<.001[b]
	Residual	2.027E+11	134458	1507261.218		
	Total	2.081E+11	134459			
2	Regression	6583996517	2	3291998258	2196.647	<.001[c]
	Residual	2.015E+11	134457	1498646.747		
	Total	2.081E+11	134459			
3	Regression	7097653029	3	2365884343	1582.703	<.001[d]
	Residual	2.010E+11	134456	1494837.635		
	Total	2.081E+11	134459			
4	Regression	7475967953	4	1868991988	1252.646	<.001[e]
	Residual	2.006E+11	134455	1492035.061		
	Total	2.081E+11	134459			
5	Regression	7725423021	5	1545084604	1036.837	<.001[f]
	Residual	2.004E+11	134454	1490190.839		
	Total	2.081E+11	134459			
6	Regression	7838616259	6	1306436043	877.179	<.001[g]
	Residual	2.002E+11	134453	1489360.043		
	Total	2.081E+11	134459			
7	Regression	7908969851	7	1129852836	758.877	<.001[h]
	Residual	2.002E+11	134452	1488847.859		
	Total	2.081E+11	134459			
8	Regression	7920922307	8	990115288	665.056	<.001[i]
	Residual	2.002E+11	134451	1488770.034		
	Total	2.081E+11	134459			
9	Regression	7953726376	9	883747375	593.702	<.001[j]
	Residual	2.001E+11	134450	1488537.120		
	Total	2.081E+11	134459			
10	Regression	7962470448	10	796247045	534.939	<.001[k]
	Residual	2.001E+11	134449	1488483.155		
	Total	2.081E+11	134459			

a. Dependent Variable: Covid.Cases

Figure 20: Model summaries of LR models predicting Covid-19 cases

Model Summary[k]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.161[a]	.026	.026	1227.706	
2	.178[b]	.032	.032	1224.192	
3	.185[c]	.034	.034	1222.636	
4	.190[d]	.036	.036	1221.489	
5	.193[e]	.037	.037	1220.734	
6	.194[f]	.038	.038	1220.393	

Figure 21: Final LR model for predicting Covid-19 cases

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-89.909	6.657		-13.506	<.001
	NO2_1st.Max.Value	17.011	.284	.161	59.989	<.001
2	(Constant)	-120.863	6.731		-17.957	<.001
	NO2_1st.Max.Value	11.843	.338	.112	35.006	<.001
	CO_1st.Max.Value	268.378	9.647	.089	27.819	<.001
3	(Constant)	77.678	12.645		6.143	<.001
	NO2_1st.Max.Value	12.181	.338	.116	35.999	<.001
	CO_1st.Max.Value	241.943	9.740	.081	24.840	<.001
	Ozone_Arithmetic.Mean	-6302.683	340.006	-.050	-18.537	<.001
4	(Constant)	138.385	13.196		10.487	<.001
	NO2_1st.Max.Value	12.025	.338	.114	35.555	<.001
	CO_1st.Max.Value	401.574	13.971	.134	28.743	<.001
	Ozone_Arithmetic.Mean	-6600.747	340.202	-.053	-19.402	<.001
	CO_Arithmetic.Mean	-504.498	31.683	-.068	-15.923	<.001
5	(Constant)	124.149	13.234		9.381	<.001
	NO2_1st.Max.Value	10.851	.350	.103	31.004	<.001
	CO_1st.Max.Value	370.767	14.164	.123	26.177	<.001
	Ozone_Arithmetic.Mean	-14140.716	674.693	-.113	-20.959	<.001
	CO_Arithmetic.Mean	-484.956	31.699	-.065	-15.299	<.001
	Ozone_1st.Max.Value	6798.357	525.447	.070	12.938	<.001
6	(Constant)	139.842	13.352		10.474	<.001
	NO2_1st.Max.Value	10.841	.350	.103	30.985	<.001
	CO_1st.Max.Value	380.003	14.200	.126	26.762	<.001
	Ozone_Arithmetic.Mean	-14108.300	674.515	-.112	-20.916	<.001
	CO_Arithmetic.Mean	-468.453	31.747	-.063	-14.756	<.001
	Ozone_1st.Max.Value	6709.412	525.399	.069	12.770	<.001
	SO2_Arithmetic.Mean	-48.353	5.546	-.024	-8.718	<.001

Time Series Analysis

Concerning COVID-19 fatalities, Sulfur Dioxide, Nitrogen Dioxide, and Carbon Monoxide exhibited the most substantial correlation values. Notably, the highest correlations with COVID-19 deaths were observed when the individual pollution series were lagged by two weeks. We extended the analysis to include up to a 15-day lag to encompass both a half-month period and the typical two-week incubation timeframe for the disease. In interpreting the lag plots, it's essential to understand that they reflect the impact of pollution on subsequent COVID-19 deaths after a specific lag period. The results from this analysis reveal two key takeaways: First, although the correlations were highest for these three pollutants, there is no distinctive lag period where correlation values significantly stand out or guide further analysis. Correlation values consistently fall within the range of 0.2 to 0.4. This suggests that while certain pollutants may have a more pronounced impact on mortality than others, pollution alone does not exert a significant influence on COVID-19 fatality rates. If pollution were a dominant factor in mortality, one would anticipate a noticeable spike in correlation at a specific lag period, signifying that elevated pollution levels on a given day result in increased deaths on subsequent days. The absence of such a pattern indicates that other factors play a more prominent role in COVID-19 mortality.

We also conducted lag plot analyses for COVID-19 transmission and case counts. In examining transmission, Sulfur Dioxide, Nitrogen Dioxide, and Carbon Monoxide, akin to our findings in the death analysis, exhibited the highest correlations when pollution levels were lagged by two weeks. These correlations remained fairly uniform within the lag analysis. If pollution levels on a specific day had a direct and predictive impact on cases occurring on a subsequent day, one would expect to observe an increase in correlation values in the lag analysis. Since this pattern does not manifest, we conclude that while pollution may not expedite the recovery from COVID-19, it is not the predominant determinant of transmission or death counts.

Overall, the findings from the lag analysis did not provide a compelling rationale for delving deeper into the intricacies of how pollution on one day directly influences deaths or transmission on later days. The complex interplay of various factors affecting COVID-19 spread, including vaccination efforts, social distancing, and improved treatment protocols, made it challenging to create a predictive model for pollution's impact on COVID-19. Moreover, pollution's effects are typically observed over years of exposure, and with the limited available data, predicting its precise influence on individuals diagnosed with COVID-19 remains challenging.

CONCLUSION

Looking ahead, it is imperative to implement measures aimed at reducing pollution levels to combat respiratory diseases effectively. While immediate pollution reduction during disease outbreaks like COVID-19 may yield modest short-term benefits, the focus should be on the long-term respiratory health of individuals residing in cleaner environments, which can potentially impact death counts and transmission rates during respiratory virus outbreaks. A notable 2012 study on large-scale ozone reduction in the United States demonstrated that even small adjustments in regulated ozone levels, such as lowering them from 75 ppb to 60 ppb, could lead to a substantial reduction in mortality. The visual representation in Figure 22 underscores the improvements achievable with stricter regulations.

Moreover, future research could explore additional time series analyses, including longer lags and the incorporation of various variables. A study addressing the impact of air pollution on COVID-19 infection and mortality (Ali, Islam, 2020) highlights the significance of long-term exposure, suggesting that lagging the time series by two months or more could provide valuable insights into pollution's influence on recovery from COVID-19 and other respiratory illnesses. Regional analyses might reveal varying impacts on areas with different pollution levels.

Our regression analysis and data discussions reveal that our linear regression model, utilizing pollutant levels as predictors for COVID-19 deaths and cases, had limited success. It accounted for less than 4% of the dataset's variation. This underscores the necessity of considering other, more influential factors in predicting COVID-19 outcomes.

These findings open opportunities for model improvement. Additional variables such as age, weight, ethnicity, health conditions, healthcare access, socioeconomic status, population density, geographic location, and vaccination status should be considered. Expanding the dataset with these factors can enhance the model's comprehensiveness. Exploring alternative machine learning models may uncover nonlinear relationships between pollutants and COVID-19 outcomes. Collaborative research efforts with experts from various disciplines can contribute to a more accurate and holistic understanding of the factors influencing COVID-19 outcomes.

Figure 22: Improvement levels in ozone

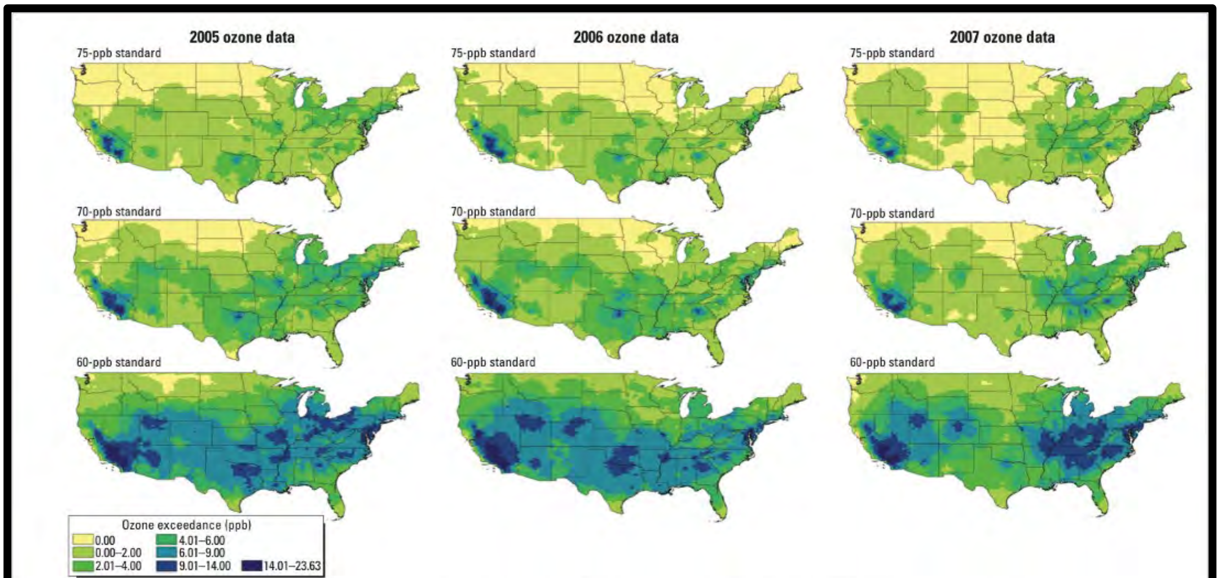


Figure 1. Estimated reductions in annual ozone (8-hr maximum) if regulatory attainments of 75, 70, and 60 ppb had been achieved (2005–2007).

REFERENCES

Al Huraimel K, Alhosani M, Kunhabdulla S, Stietiya MH. SARS-CoV-2 in the environment: modes of transmission, early detection and potential role of pollution. Sci Total Environ. (2020) 744:140946. doi: 10.1016/j.scitotenv.2020.140946

Ali N., Islam F. 2020. The Effects of Air Pollution on COVID-19 Infection and Mortality—A Review on Recent Evidence. Frontiers In Public Health Section Environmental health and Exposome Volume 8 - 2020

Anenberg SC, Miller J, Minjares R, Du L, Henze DK, Lacey F, et al.2017. Impacts and mitigation of excess diesel-related NOx emissions in 11 major vehicle markets. Nature545(7655):467–471, PMID: 28505629, 10.1038/nature22086.

Berman JD, Fann N, Hollingsworth JW, Pinkerton KE, Rom WN, Szema AM, Breysse PN, White RH, Curriero FC. Health benefits from large-scale ozone reduction in the United States. Environ Health Perspect. 2012 Oct;120(10):1404-10. doi: 10.1289/ehp.1104851. Epub 2012 Jul 18. PMID: 22809899; PMCID: PMC3491929.

Brauer M. How much, how long, what, and where: air pollution exposure assessment for epidemiologic studies of respiratory disease. Proc Am Thoracic Soc. (2010) 7:111–5. doi: 10.1513/pats.200908-093RM

Briggs DJ, Sabel CE, Lee K. 2009. Uncertainty in epidemiology and health risk and impact assessment. Environ Geochem Health 31:189–203

Centers for Disease Control and Prevention, 2020. COVIDView: A Weekly Surveillance Summary of US. COVID-19 Activity [WWW Document]. URL https://www.cdc.gov/coronavirus/2019-ncov/covid-data/pdf/covidview-07-24-2020.pdf

Comunian S, Dongo D, Milani C, Palestini P. Air pollution and COVID-19: the role of particulate matter in the spread and increase of COVID-19's morbidity and mortality. IJERPH. (2020) 17:4487. doi: 10.3390/ijerph17124487

Copat C, Cristaldi A, Fiore M, Grasso A, Zuccarello P, Oliveri Conti G, et al. A first review to explore the association of air pollution (PM and NO2) on Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV-2). Med Pharmacol. (2020). doi: 10.20944/preprints202005.0299.v1

Department of the Environment, Water, Heritage and the Arts. (n.d.). Carbon Monoxide. Australian Government . https://www.dcceew.gov.au/environment/protection/npi/resource/student/carbon-monoxide#:~:text=When%20carbon%20monoxide%20is%20emitted,causing%20other%20extreme%20weather%20events

DeVilbiss, J., & Ray, S. (2017, February 3). Sulfur dioxide emissions from U.S. power plants have fallen faster than coal generation. U.S. Energy Information Administration (EIA). https://www.eia.gov/todayinenergy/detail.php?id=29812

Domingo JL, Marquès M, Rovira J. Influence of airborne transmission of SARS-CoV-2 on COVID-19 pandemic. A review. Environ Res. (2020) 188:109861. doi: 10.1016/j.envres.2020.109861

E.B. Brandt, A.F. Beck, T.B. Mersha. Air pollution, racial disparities, and COVID-19 mortality J. Allergy Clin. Immunol. (2020), 10.1016/j.jaci.2020.04.035

Environmental Protection Agency. (2021, October 7). Basic Ozone Layer Science. EPA. https://www.epa.gov/ozone-layer-protection/basic-ozone-layer-science

Environmental Protection Agency. (2023b, March 15). NAAQS Table. EPA. https://www.epa.gov/criteria-air-pollutants/naaqs-table

Environmental Protection Agency. (2023a, February 16). Sulfur Dioxide Basics. EPA. https://www.epa.gov/so2-pollution/sulfur-dioxide-basics#:~:text=What%20are%20the%20environmental%20effects,which%20can%20harm%20sensitive%20ecosystems.

GBD. Global, regional, and national comparative risk assessment of 84 behavioral, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990−2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Lond Engl. (2018) 392:1923–94.

Kim, A. (2022, February 16). US pollution 2000-2021. Kaggle. https://www.kaggle.com/datasets/alpacanonymous/us-pollution-20002021

Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air pollution sources to premature mortality on a global scale. Nature. (2015) 525:367–71. doi: 10.1038/nature15371

Lipsitt et al., Spatial analysis of COVID-19 and traffic-related air pollution in Los Angeles, Environment International, Volume 153, 2021, 106531, ISSN 0160-4120, https://doi.org/10.1016/j.envint.2021.106531.

Messier KP, Chambliss SE, Gani S, Alvarez R, Brauer M, Choi JJ, et al.2018. Mapping air pollution with Google Street View cars: efficient approaches with mobile monitoring and land use regression. Environ Sci Technol 52(21):12563–12572, PMID: 30354135, 10.1021/acs.est.8b03395. Crossref, Medline, Google Scholar

U.S. Department of Health & Human Services. (2021, November 25). Air Quality Measures on the National Environmental Health Tracking Network. Data.Gov. https://catalog.data.gov/dataset/air-quality-measures-on-the-national-environmental-health-tracking-network

World Health Organization. (2021, October 30). Climate Change and Health. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health

World Health Organization
WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard [WWW Document] https://covid19.who.int/

Zhou et al. Excess of COVID-19 cases and deaths due to fine particulate matter exposure during the 2020 wildfires in the United States. Sci. Adv.7, eabi8789 (2021). DOI:10.1126/sciadv.abi8789.

DECISION SCIENCES INSTITUTE

Forecasting the Dynamic Demand-Supply Gap by an LSTM-based Recurrent Neural Network Model in the Chicago City Taxi and Ridesharing Market: A Post-Pandemic Analysis

Ilyas Ustun
DePaul University
iustun@depaul.edu

Zhong Dong
DePaul University
ZDONG6@depaul.edu

ABSTRACT

This paper presents a post-pandemic analysis of the Chicago City taxi and ridesharing market and utilizes an LSTM-based recurrent neural network model to forecast the dynamic demand-supply gap in the O'Hare International Airport and Chicago downtown area. Training the model by trip data and additional holiday and weather data from January 2022 to November 2022, and testing on data from December 2022, the model achieved promising results. The analysis reveals market segmentation and significant temporospatial patterns of demand-supply gap. The results provide insights for policymakers, city administrators, and industry professionals seeking to optimize taxi and ridesharing operations using deep learning techniques.

KEYWORDS: Taxi, Ridesharing, Post-Pandemic Analysis, Demand-Supply Gap, Forecasting, Deep Learning, Recurrent Neural Network, LSTM, Chicago City, Urban Transportation

INTRODUCTION

Taxis have been a crucial component of urban transportation for many years. Accurate forecasting of taxi demand and supply is critical to City Intelligent Transportation Systems, which aims to reduce customer waiting time, enhance the efficiency of taxi spatiotemporal operations, alleviate city traffic congestion, and reduce vehicle emissions. In recent years, Transportation Network Companies (TNCs), such as Uber and Lyft, have emerged and grown exponentially, competing with traditional taxis, and changing the face of the industry. In this paper, we use the term "taxi" to refer to the traditional taxis and "ridesharing" to refer to the vehicle services provided by TNCs.

The taxi and ridesharing industry have experienced significant changes due to the emergence of new technologies and the impact of the COVID-19 pandemic. The pandemic has created new challenges for urban transportation, including changes in travel behavior, reduced demand for taxi and ridesharing services, and shifts in travel patterns. Addressing these challenges necessitates a thorough understanding of the current state of the taxi and ridesharing market, as well as the development of new strategies to enhance operational efficiency and competitiveness.

This paper presents a post-pandemic analysis of the Chicago City taxi and ridesharing market from January 2019 to December 2022. We focus on forecasting the dynamic demand-supply gap using an LSTM-based recurrent neural network model. The study utilizes trip data from January to November 2022, as well as additional holiday and weather data, to identify the temporospatial patterns of demand and supply in the O'Hare International Airport and Chicago downtown area.

We sourced our data from (Chicago Data Portal, 2023) , which provides taxi trip data for the City of Chicago from January 2013 onwards, updated monthly. It also provides trip data reported by ridesharing companies such as Uber and Lyft from November 2018 onwards, updated monthly. We also use the hourly weather data at the O'Hare International Airport from (OpenWeather, 2023).

The definition of demand and supply can vary in different studies. In this paper, we define the demand and supply for taxis or ridesharing as follows: in a given area for each half-hour period, the number of taxis or ridesharing vehicles picking up customers is defined as demand, assuming there are no customer waiting times, and the pickup quantity equals the customer demand. The number of taxis or ridesharing vehicles dropping off customers is defined as supply, assuming that these vehicles are available to pick up new customers in the same area. The demand-supply gap is defined as the difference between demand and supply in that time and area.

By analyzing the historical data on demand and supply and developing a deep learning model to forecast the demand-supply gap, our study aims to reveal the net supply surplus or shortage in a specific area and time period of the city.

Specifically, this paper will address the following four research questions:

1. What are the market trends and segmentations for taxis and ridesharing in Chicago throughout the COVID-19 pandemic?
2. What is the current status of taxi operational efficiency?
3. Are there any temporospatial patterns for the demand-supply gap in taxis and ridesharing services?
4. How can we effectively utilize deep learning to forecast the demand-supply gap?

The remaining sections of this paper are organized as follows: In the Literature Review section, we introduce related works on taxi and rideshare demand predictions that use machine learning or deep learning. The Exploratory Data Analysis section presents insights into the taxi and ridesharing market based on the analysis of trip data. In the Demand-Supply Gap Prediction Model section, we describe the data processing and model training method used to forecast the dynamic demand-supply gap by an LSTM-based recurrent neural network model. The Results section evaluates the model test results. Finally, in the Conclusions and Discussion section, we summarize the main findings of this study, provide insights into the potential implications of our research, and discuss future work.

LITERATURE REVIEW

The temporospatial dynamics of traffic are of significant interest in transportation studies. Demand prediction is a crucial aspect within this domain. Early demand predictions were widely using autoregressive integrated moving average model (ARIMA) from time series analysis to forecast future demand. Other machine learning models such as linear regression, support vector machine (SVM), k-nearest neighbor (KNN), were also utilized.

In recent years, deep learning models based on neural networks (NNs) have been applied in this field. Among them, long short-term memory (LSTM) based recurrent neural network (RNNs) exhibited promising results(Ke et al., 2017; Zhao et al., 2017). In the related studies, the data source, the method, and the definition of demand and supply vary.

(Li & Allan, 2020) conducted a study on ridesharing in Chicago citywide and proposed a taxi efficiency optimization solution based on demand prediction. They defined the supply-demand gap as the number of available rideshare vehicles compared to the demand learned from the rideshare company's rider request records.

(Chebance et al., 2021) performed simulations on Chicago Lincoln Park and O'Hare Airport ridesharing data and compared the performance of three supply-demand matching policies. They defined demand as individual rider requests, and supply as the to-be-matched vehicles.

In comparison to ridesharing, most taxi services primarily rely on actively searching for passengers on the road. This means that they do not have a centralized online request record or information about idle taxis readily available. (Li et al., 2022) defined taxi demand as actual pickup numbers in every 15 minutes. They utilized taxi trip data in Chicago and proposed a taxi trip optimization strategy in the downtown area with demand prediction using a LSTM model.

(Wali et al., 2022) developed a ridesharing demand model based on factors such as Chicago census tracts' environment, transit accessibility, crime, active travel, and demographic factors. They defined demand as ridesharing trip quantity per 100 individuals in a specific period.

(Faial et al., 2020) used New York City (NYC) data and proposed a neural network model for taxi demand prediction using stream learning. They defined demand as the number of vehicles which passengers have already boarded, disregarding others possibly waiting in line. This assumption aligns with our study.

(Qian et al., 2020) utilized NYC Manhattan area rideshare trip data and applied the Gaussian Conditional Random Field (GCRF) model to forecast ridesharing short-term demand. The demand was also defined as the actual pickup taxi quantity in a specific period.

Apart from solely modeling demand, some studies also involve additional factors. (Dey et al., 2021) utilized NYC ridesharing trip records and presented a model for the ridesharing demand and flow distribution patterns. The model considered impact factors such as the built environment, socio-demographic characteristics, and weather information.

(Zhang et al., 2021) proposed a taxi demand prediction model called Multi-Level Recurrent Neural Networks (MLRNN), clustering NYC taxi zones into several clusters and combining the cluster-level and the global-level prediction modules. Hourly weather data in NYC was used in the model.

In addition to studying demand, some studies also explore supply. (Hao et al., 2020) utilized Singapore taxi operator data and defined the idle vehicles as supply, while passengers who failed to find a vehicle via mobile app were considered as demand. They used weather factors (precipitation and rainfall time) as covariates for unsupervised machine learning.

(Wang et al., 2017) introduced the concept of rideshare supply-demand gap and presented a deep neural network to predict the gap between supply and demand. They used orders data from Didi, the biggest rideshare company in China, and defined the supply-demand gap as the total amount of invalid orders in an area and within a time interval. Their model also involved weather data and area traffic data.

In these studies, the supply data usually come from taxi or ridesharing companies. Our study is based on available public data, where we define the demand and supply as the real pickup and drop-off numbers within a specific area and time frame. Our focus is specifically on analyzing the gap between demand and supply. Additionally, considering that weather conditions and holidays significantly impact taxi and ridesharing demand and supply, we incorporate these factors into our prediction model.

DATA

(Chicago Data Portal, 2023) offers data of all taxi trips starting from January 2013. In the dataset, each row represents an individual trip, and has 23 columns including trip ID, taxi ID, trip start timestamp, trip end timestamp, trip seconds, trip miles, pickup census tracts, drop-off census tract, pickup community area, drop-off community area, fare, tips, tolls, extras, trip total, etc. Census tracts are suppressed in some cases, and all times are rounded to the nearest 15-minute interval.

(Chicago Data Portal, 2023) also provides data of all trips reported by ridesharing companies since November 2018. Similar to the taxi trip dataset, each row represents a trip and has 21 columns including trip ID, trip start timestamp, trip end timestamp, trip seconds, trip miles, pickup census tracts, drop-off census tract, pickup community area, drop-off community area, fare, tip, additional charges, trip total, shared trip authorized, trips pooled, etc. Census tracts are suppressed in some cases, and all times are rounded to the nearest 15-minute interval. Fares are rounded to the nearest $2.50 and tips are rounded to the nearest $1.00. To protect privacy, there is no vehicle IDs.

Taxi and Ridesharing Trip Quantity from 2019 to 2022

We aggregated the daily trip quantities from the start of 2019 until the end of 2022. Figure 1 illustrates the trend in daily trip quantities for both taxi and ridesharing services since 2019. We can see prior to the COVID-19 pandemic, ridesharing has already dominated the market. Then the pandemic led to a sharp drop down in trip quantities starting from March 20, 2020, coinciding with the implementation of the Illinois statewide stay-at home order. Since then, both taxi and ridesharing trip quantities have experienced slow and gradual recovery, but the ridesharing demonstrated a faster recovery.

Figure 1: Taxi and rideshare daily trip quantity in Chicago (2019-2022)

Figure 2 and Figure 3 depict the monthly trip quantities for taxis and ridesharing respectively over the four-year period. Comparing the data for 2019 and 2022, it is evident that the monthly trip quantity for taxis in December 2022 amounted to only 41.8% of that recorded in December 2019. Conversely, the ridesharing monthly trip quantity had recovered to 67.6% of the December 2019 amount.

Figure 2: Taxi monthly trip quantity in Chicago (2019-2022)

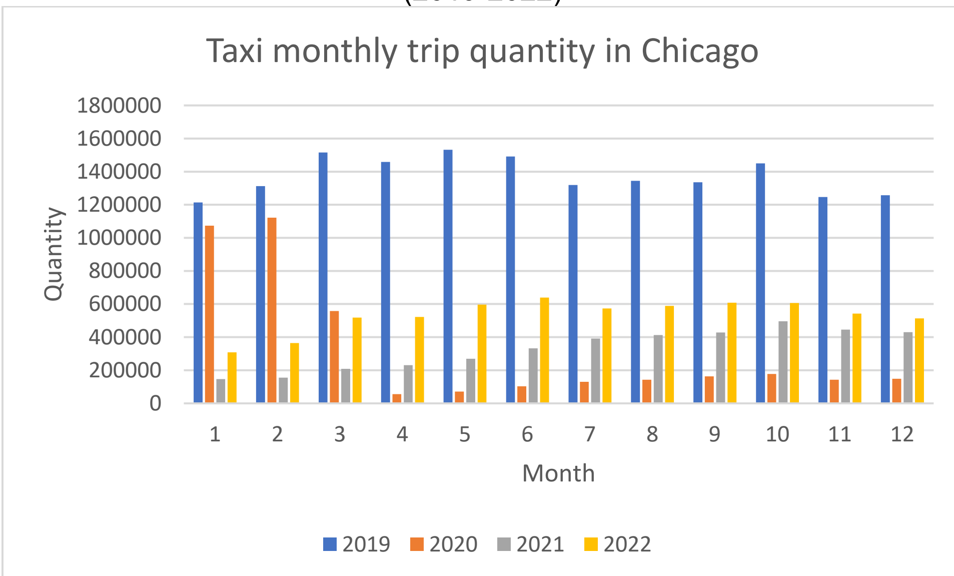
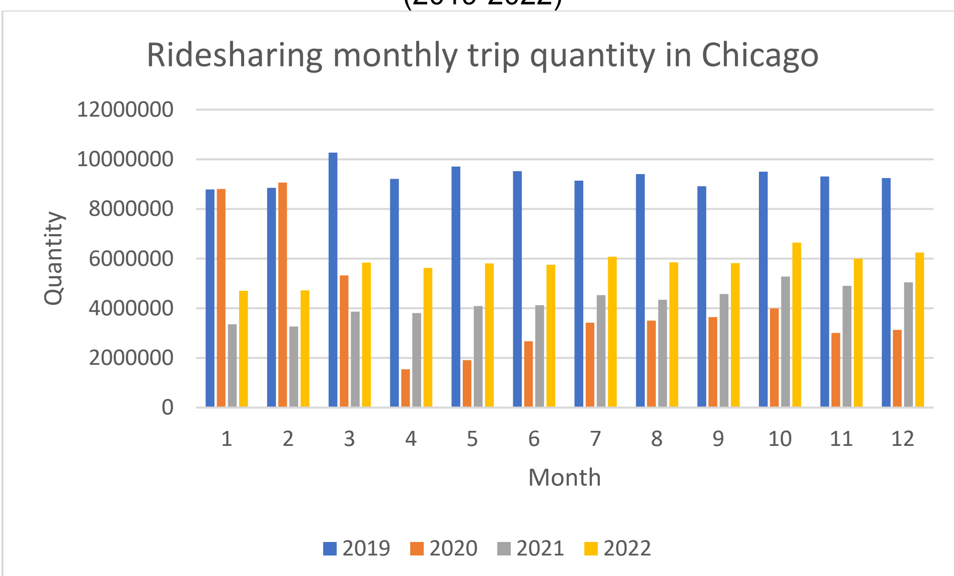
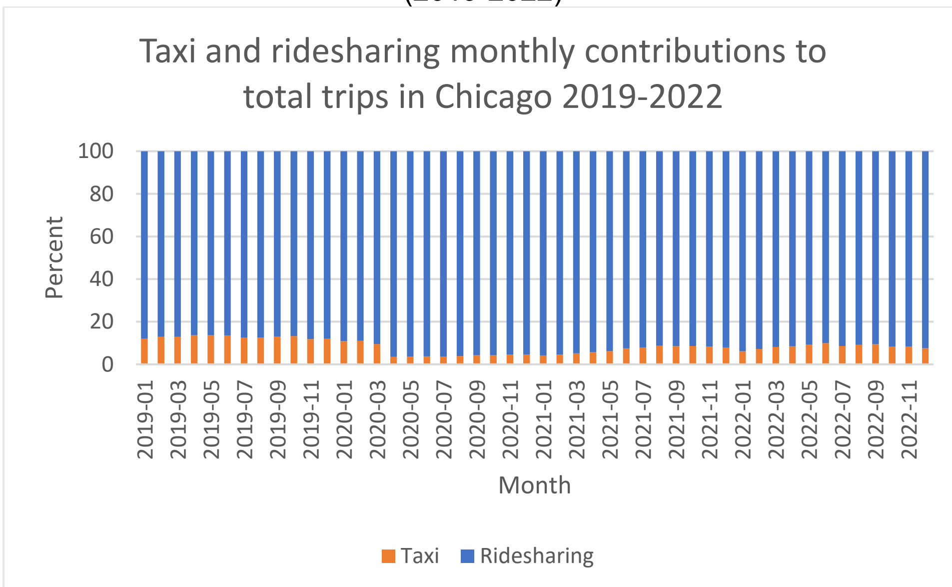


Figure 3: Ridesharing monthly trip quantity in Chicago (2019-2022)



As shown in Figure 4, the market share of taxis experienced a significant decline at the beginning of the pandemic, followed by a gradual ascent since 2021. Nevertheless, the taxi market remains in a state of shrinkage when compared to the pre-pandemic conditions of 2019.

Figure 4: Taxi and ridesharing monthly contributions to total trips in Chicago (2019-2022)



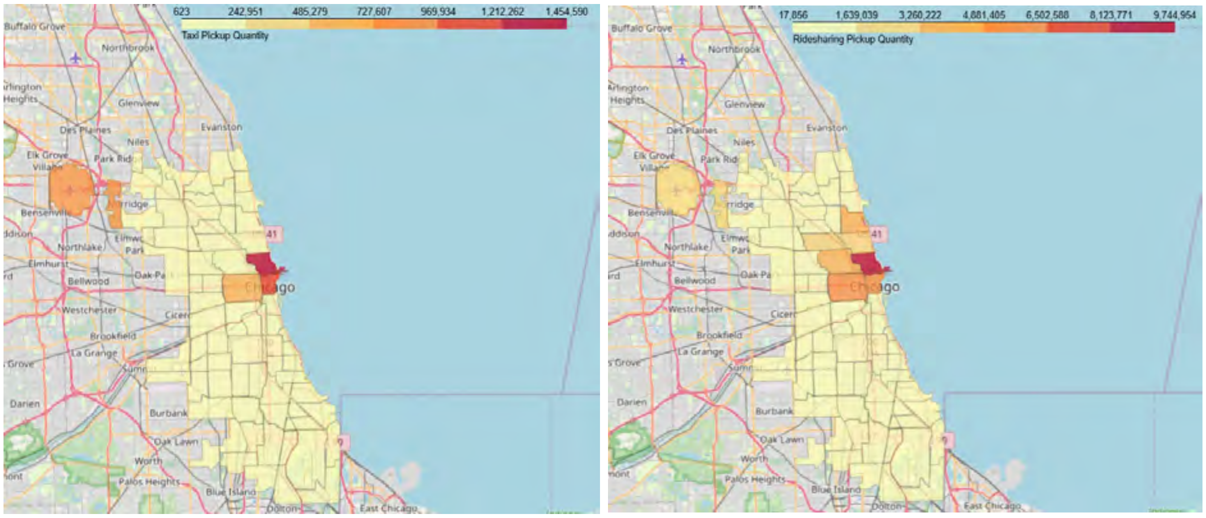
Taxi and Ridesharing Market Segmentation in 2022

To study taxi and ridesharing market segmentation in post-pandemic, we analyzed the trip data from 2022, which are 6,382,425 taxi trips and 69,081,663 ridesharing trips.

Each trip includes the pickup and drop-off community area numbers. The Chicago City has 77 community areas(Chicago Data Portal, 2023). For example, community 76 is O'Hare International Airport, community 8 is 'Near Northside', community 32 is 'Loop' (Downtown), etc. For the pickup or drop-off area outside the city, the community area number is "NA". To focus our study on the city, trips with "NA" areas were removed, resulting in 5,471,412 taxi trips and 57,295,942 ridesharing trips, accounting for 85.7% and 82.9% of the total, respectively.

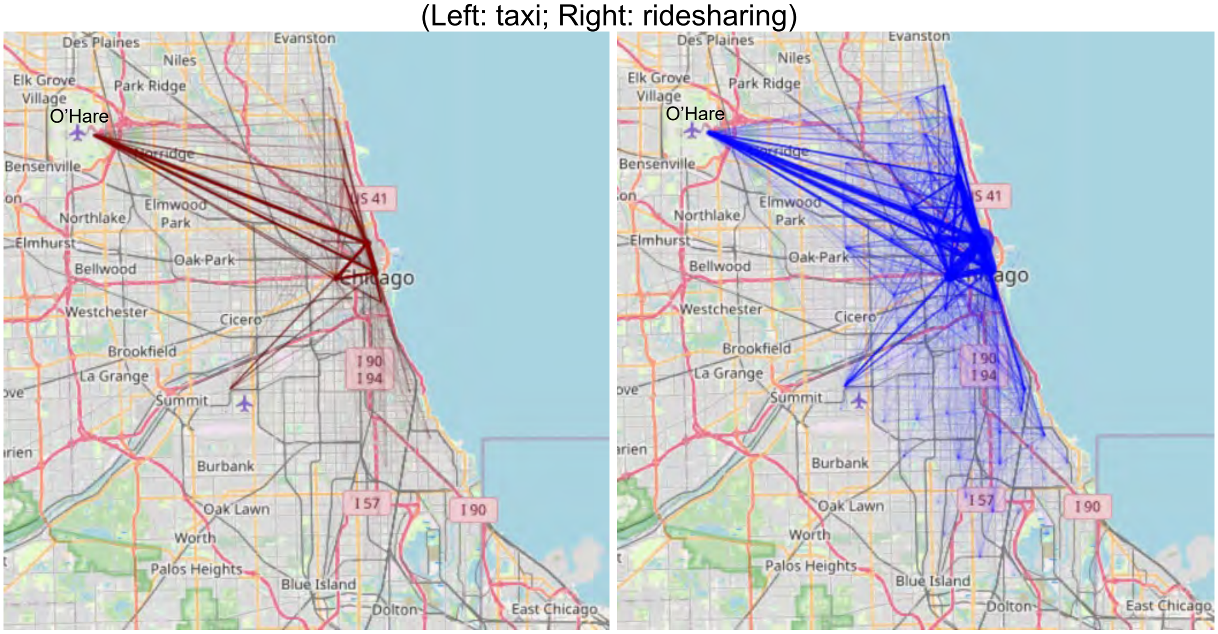
Figure 5 illustrates the distribution of pickups in the 77 community areas of Chicago, with the left side showing taxi pickups and the right side showing ridesharing pickups. The heat bar marks indicate that ridesharing pickups are significantly higher than taxi pickups. Both taxi and ridesharing pickups are concentrated in downtown and airport areas, but ridesharing pickups extend further to the west and north of downtown.

Figure 5: Taxi and ridesharing pickup quantity in Chicago (2022)
(Left: taxi; Right: ridesharing)



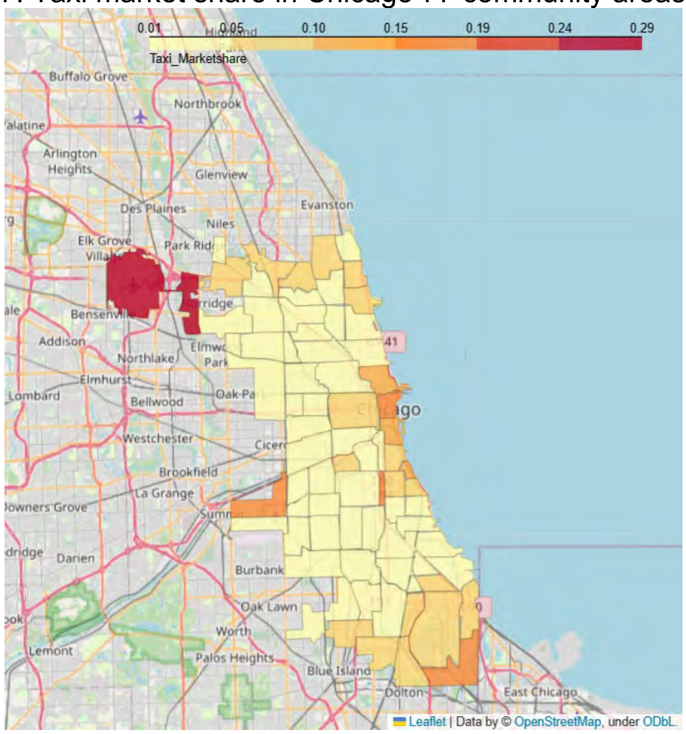
To analyze the density of trip origins (O) and destinations (D), we aggregated the quantity of OD pairs between community areas. Figure 6 depicts the origin and destination as the geographic center of the pickup or drop-off community area. The red lines in the left figure represent taxi OD pairs, while the blue lines in the right figure represent ridesharing OD pairs. The width of the lines indicates the density of OD pairs: wider lines represent more pairs. We found taxi OD pairs are highly concentrated within the downtown area and between downtown and the two airports, O'Hare International Airport and Midway International Airport. Ridesharing OD pairs connect a wider range of areas, but downtown and the airports still contribute significantly.

Figure 6: Taxi and ridesharing OD pairs in Chicago (2022)
(Left: taxi; Right: ridesharing)



Furthermore, we calculated the rate of taxi pickups compared to the total pickups of both taxis and ridesharing vehicles, representing the taxi market share based on pickup quantity. Figure 7 displays the taxi market share in different community areas. The highest share is observed at O'Hare International Airport (28.8%), followed by Midway Airport (18.7%) and the Loop (16.7%).

Figure 7: Taxi market share in Chicago 77 community areas (2022)



Taxi Trajectory and Efficiency at O'Hare International Airport

The taxi trip dataset includes the taxi ID for each trip, enabling us to study the taxi trajectories and efficiency. As O'Hare International Airport has the highest market share among other areas, we filtered out taxi trips in 2022 where the pickup area was identified as 76 (O'Hare International Airport), resulting in a total of 989,907 trips.

By examining the previous trip of each taxi (trip n-1), we identified the drop-off area before the taxi proceeded to the airport. We found that only 5.8% of taxis dropped off and then picked up customers at the airport, while the remaining taxis originated from different areas and arrived at the airport without any passengers. Figure 7 illustrates the percentage of top origins. The numbers around the pie chart represent community area numbers. The top origins include community area 8 (Near North Side) with 26.3%, community area 32 (Loop) with 16.3%, community area 28 (Near West Side) with 9.4%, community area 76 (O'Hare) with 5.8%. Figure 8 shows the connecting lines of vacant taxis driving to the airport, with the line width representing the quantity of the taxis. This raises the problem of a high rate of vacant taxis driving to the airport, resulting in a loss of resources, and impacting the revenue of taxi companies and drivers, as well as increasing congestion and the total vehicle emissions in the city.

Since the ridesharing trips data have no vehicle IDs, we cannot track the vehicle's trip before the airport picking up. Thus, we cannot perform the similar analysis for ridesharing with current available data.

Figure 7: Top origin area before airport pickup

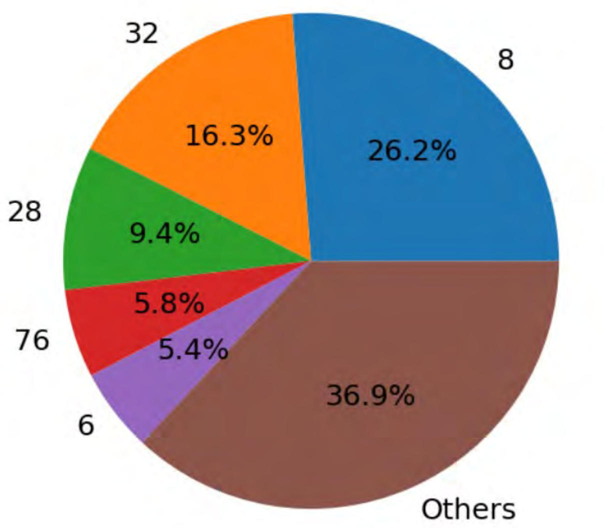
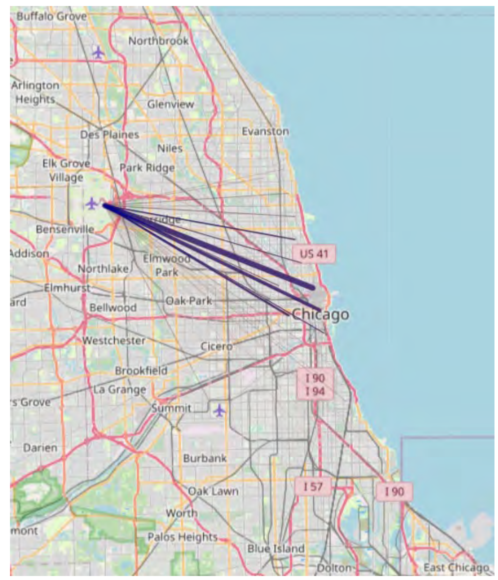


Figure 8: Vacant taxis driving to the airport



Taxi and Ridesharing Demand-Supply Gap Temporospatial Patterns

Building on the interest in the taxi and ridesharing flow and efficiency in certain area, our research is focused on examining the dynamic gap between the demand and supply. As outlined in the introduction section, we define demand as the number of pickups, supply as the number of drop-offs, and the difference between the two as the demand-supply gap for each half-hour period in a given area.

We choose O'Hare International Airport and Chicago downtown as study focus, as these areas are the top markets for both taxis and ridesharing. In this paper, O'Hare International Airport is referred to as community 76, while downtown area is referred to as community 8 (Near North Side) and community 32 (Loop).

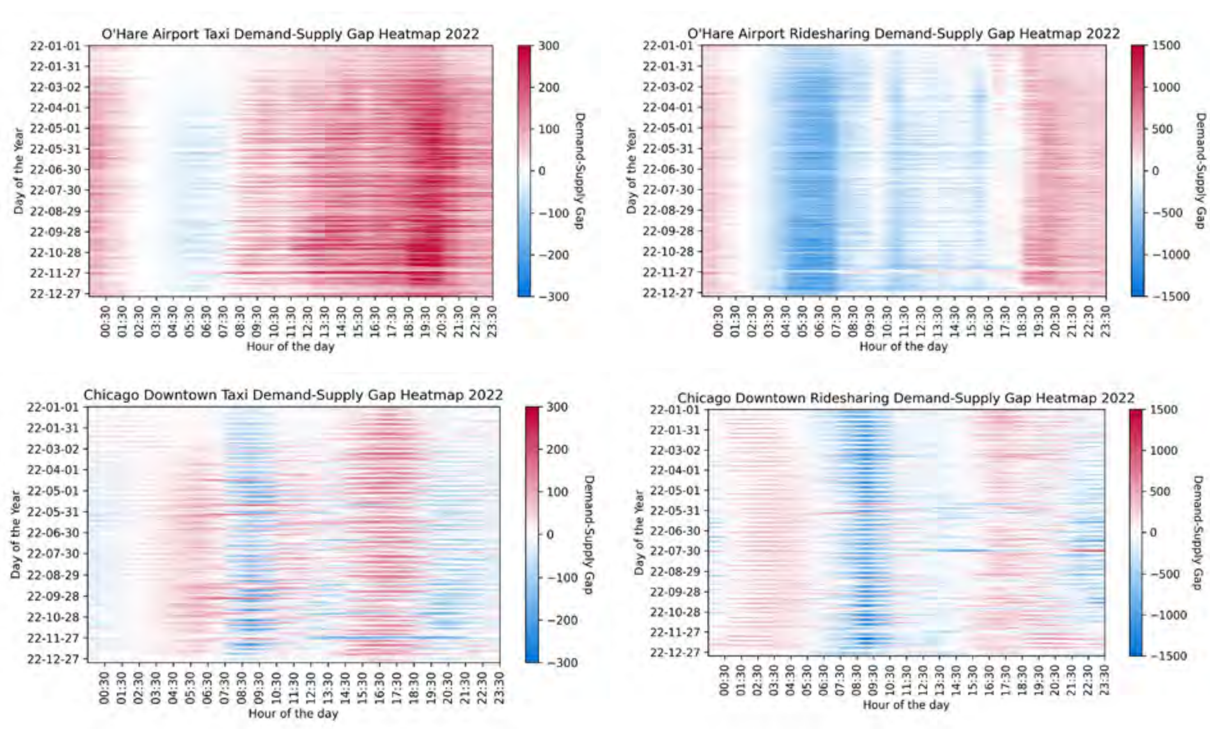
Upon investigating the demand-supply gap for taxis and ridesharing services at O'Hare International Airport and the downtown area over the course of one year, we discovered distinctive daily patterns. Figure 9 showcases the heatmaps representing the demand-supply gap for taxis (left) and ridesharing (right). The X-axis corresponds to half-hour intervals throughout the day, while the Y-axis represents the day of the year 2022. The red color indicates a positive demand-supply gap, suggesting that there were more vacant taxis or ridesharing vehicles arriving from other areas (net inflow). Conversely, the blue color represents a negative demand-supply gap, indicating that vacant taxis or ridesharing vehicles were departing to other areas (net outflow).

At the airport, the taxi demand-supply gap is mostly positive, except during the early morning hours from 3:30 am to 7:30 am, indicating that pickups exceed drop-offs for the majority of the day. This aligns with our findings, indicating a higher number of vacant taxis driving to the airport than occupied taxis. On the other hand, the ridesharing demand-supply gap is negative during the morning and noon hours, transitioning to a positive gap after 14:30 and extending until 1:30 am the following day.

In the downtown area, both taxis and ridesharing exhibit similar peak hours. From 8:30 am to 9:30 am, there is a negative peak, indicating more drop-offs than pickups at the start of the workday. For taxis, the positive peak occurs from 15:30 to 17:30, representing a greater number of pickups than drop-offs. Ridesharing experiences a positive peak starting at 16:30, extending until 21:30.

Additionally, the Y-axis of the heatmaps reveals patterns related to the day of the year. These observations inspire our further exploration into the predictability of the demand-supply gap using contemporary machine learning technologies.

Figure 9: Taxi and ridesharing pickup density at O'Hare airport and downtown area



DEMAND-SUPPLY GAP PREDICTION MODEL

In this section, we will describe our model to predict the gap between taxi demand and supply in a specific area.

Definitions

Demand: Quantity of pickups in a specific area in each half-hour interval
Supply: Quantity of drop-offs in a specific area in each half-hour interval
Demand-Supply Gap: Demand minus Supply in an area in a half-hour time interval
O'Hare International Airport area: community 76 defined in (Chicago Data Portal, 2023)
Chicago downtown area: community 8 and 32 defined in (Chicago Data Portal, 2023)

Data and Data Pre-Processing

Taxi and Ridesharing Trips Data

We use (Chicago Data Portal, 2023) taxi and ridesharing trips data from 2022. Table 1 summarized the total pickups and drop-offs in O'Hare International Airport and Chicago downtown area.

Table 1: Numbers of Pickups and Dro-offs in 2022

	Taxi		Ridesharing	
	Pickups	Drop-offs	Pickups	Drop-offs
O'Hare	989,907	352,543	3,256,472	4,045,437
Downtown	2,513,225	2,443,278	15,471,577	15,728,973

We group the pickups or drop-offs to each half hour and get the pickup and drop-off quantities in every half-hour interval.

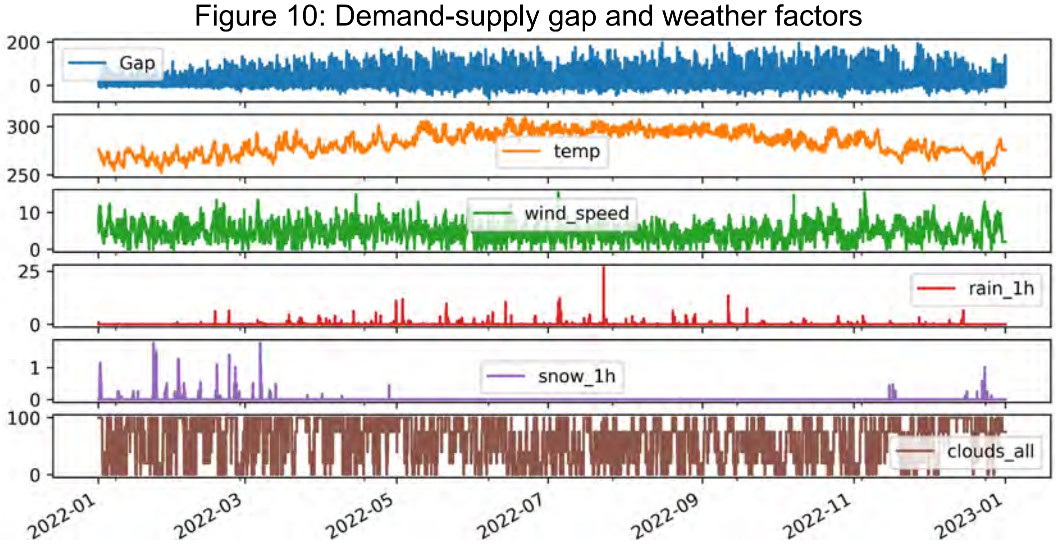
Holiday Data & Weather data

We use 2022 public holidays in Illinois and add one variable 'Holidays': 1 for holiday and 0 for non-holiday. In addition, we use hourly weather data at O'Hare International Airport from (OpenWeather, 2023).

main_temp: Temperature, Kelvin
wind_speed: Wind speed, meter/sec
rain_1h: Rain volume for the last hour, mm
snow_1h: Snow volume for the last hour, mm (in liquid state)
clouds_all: Cloudiness, %
weather_description: Weather condition within the group

Figure 10 illustrates the taxi demand-supply gap time series and the weather factors at O'Hare International Airport.

Figure 10: Demand-supply gap and weather factors



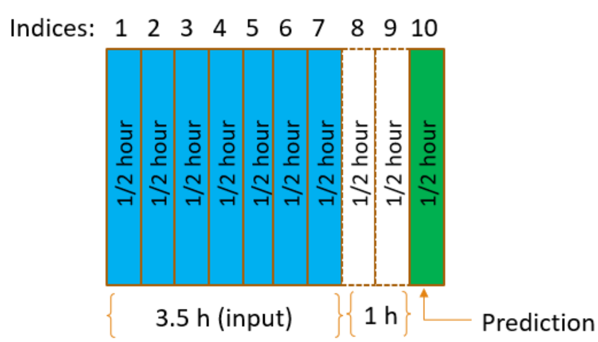
Methods and Models

Recurrent Neural Networks (RNNs) are a type of neural network designed for sequence modeling tasks, where the input data has a sequential or temporal structure. Long short-term memory (LSTM) (Sepp Hochreiter and Jürgen Schmidhuber, 1997) is a variant of RNN that addresses the vanishing gradient problem associated with traditional RNNs. It introduces memory cells and gates and enables LSTM networks to capture long-term dependencies in sequences. Bidirectional Recurrent Neural Networks (BRNNs) (M. Schuster and K. K. Paliwal, 1997) are a type of recurrent neural network that process input sequences in both forward and backward directions, taking into account both past and future context when making predictions or analyzing sequences. Gated recurrent unit (GRU) RNNs (Gulcehre, 2014) (Chung & Bengio, 2014) is another variant of RNNs that addresses the limitations of traditional RNNs. GRU combines the memory cell and the hidden state into a single unit, simplifying the architecture compared to LSTM. GRU also introduces gates, such as the update gate and the reset gate, to control the flow of information. These gates enable the GRU network to selectively update and forget information, making it computationally efficient while still capturing long-term dependencies. Autoregressive RNNs (AR-RNNs) is class of RNNs which are trained to predict the next step in a sequence given the previous steps. AR-RNNs use their own predictions as input at each time step, making them self-referential models. AR-RNNs and LSTM can be combined to create autoregressive LSTM models. In such cases, LSTMs are used as the architecture for modeling the temporal dependencies and generating autoregressive predictions.

In this paper, we use RNNs framework to build our prediction models in the TensorFlow framework (TensorFlow, 2023). We build several models using LSTM RNNs, Bidirectional RNNs, GRU RNNs, and Autoregressive LSTM RNNs.

In the model, we use seven data points to predict the tenth data point as demonstrated in Figure 11. Because each data point is for half-hour interval, our predictions are using 3.5-hour data input to predict the demand-supply after one hour.

Figure 11: Input indices and prediction indices



RESULTS

We use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to evaluate the model:

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - x_i)^2}{n}} \quad (2)$$

MAE = mean absolute error
RMSE = root mean square error
y_i = prediction
x_i = true value
n = total number of data points

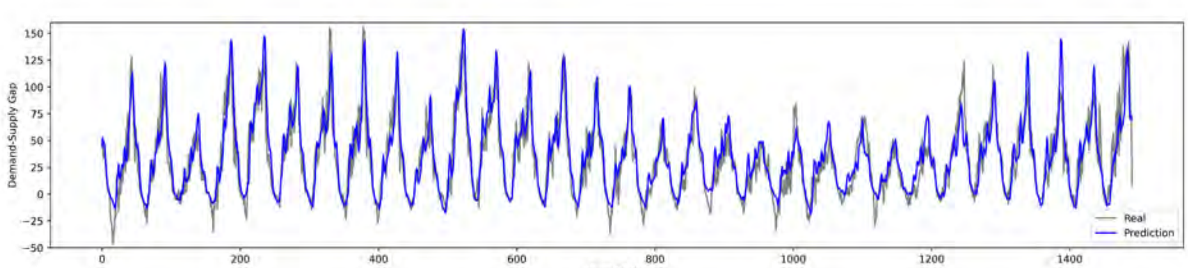
We use data from January 2021 to November 2021 to train the model. During training, we use 80% data for training and 20% data for validation. Then we test the model with data from December 2021 and compare the forecast to the real demand-supply gap. Table 2 provides a summary of our test results after de-normalizing the data. The numbers in gray shadow are the best results from four RNN models.

Table 2: Results of Demand-Supply Gap Forecasting Using Different Models

		LSTM RNNs		Bi-directional RNNs		GRU RNNs		Autoregressive LSTM RNNs	
		MAE	RMSE	MAE	RSME	MAE	RMSE	MAE	RSME
Taxi	O'Hare	11.8	16.1	14.6	19.6	15.0	20.2	11.5	15.5
	Downtown	13.5	18.3	15.3	20.3	14.1	18.9	13.2	18.0
Ride-sharing	O'Hare	53.1	72.1	55.5	73.7	57.6	76.5	55.8	74.8
	Downtown	70.6	99.0	72.0	99.8	68.1	95.2	68.2	94.6

Autoregressive LSTM models get the best results in 3 cases and typical LSTM models get the best result in one case. Figure 12 illustrates the comparison between the Taxi demand-supply gap forecast and the actual gap at O'Hare International Airport during December 2022 using Autoregressive LSTM model. The gray line represents the real data and the blue line represents the predictions. The results are promising.

Figure 12: Taxi demand-supply gap forecasting test results (O'Hare December 2022)



CONCLUSIONS AND DISCUSSIONS

The paper presented an exploratory data analysis of taxi and ridesharing trip data in Chicago, focusing on the period from 2019 to 2022. Additionally, the paper proposed a prediction model using different types of recurrent neural networks (RNNs) to forecast the demand-supply gap for taxi and ridesharing at O'Hare International Airport and Chicago downtown area.

Market Trend and Segmentation

The analysis showed that prior to the pandemic, ridesharing dominated the market, but both taxi and ridesharing trip quantities dropped significantly during the pandemic. While both sectors experienced gradual recovery, ridesharing demonstrated a faster rebound, reaching 67.6% of the pre-pandemic trip quantity by December 2022, compared to only 41.8% for taxis.

The analysis examined the distribution of pickups in the 77 community areas of Chicago. Ridesharing pickups were significantly higher than taxi pickups and extended further to the west and north of downtown. Taxi trips were concentrated within downtown and between downtown and the airports.

Taxi Efficiency at O'Hare International Airport

A high percentage of taxis drove to O'Hare International Airport without passengers, leading to a loss of resources and increased congestion and emissions. The study identified top origins for vacant taxis driving to the airport, such as the Near North Side and the Loop.

Demand-Supply Gap Temporospatial Patterns

The analysis focused on the demand-supply gap for taxis and ridesharing services at O'Hare International Airport and Chicago downtown. Distinctive daily patterns were observed at the O’Hare airport, with taxis showing a positive gap for most of the day, indicating more pickups than drop-offs. Ridesharing had a negative gap during morning and noon hours and a positive gap during the afternoon and evening.

Demand-Supply Gap Prediction Model

A predictive model was developed using Recurrent Neural Networks (RNNs). The model used taxi and ridesharing historical data on pickups, drop-offs, public holidays, and weather conditions. The Autoregressive LSTM model achieved the best performance in predicting the demand-supply gap.

Discussion and Future Works

The findings of this study suggest the need for taxi companies to adapt and innovate to remain competitive in the changing market. Understanding the market segmentation and demand-supply gap patterns can help taxi companies identify potential areas for business growth and strategic decision-making.

The analysis of taxi efficiency at O'Hare International Airport highlights the need for better coordination between taxis and city authorities. Implementing measures to reduce the number of vacant taxis driving to the airport could improve resource utilization, reduce congestion, and minimize environmental impacts. This could involve flexible fare rate and tax rate, or customer reward programs to incentivize taxis to pick up passengers on their way to the airport. The airport could offer convenient parking for taxis that have dropped off passengers, allowing them to wait for peak demand times.

The demand-supply gap patterns can assist in optimizing service availability during peak hours and improving operational efficiency. The proposed prediction model for forecasting the demand-supply gap can be further refined and integrated into real-time systems to support timely decision-making and vehicle allocation.

Our future works could be extending the analysis to more areas and providing a city-scale understanding of the taxi and ridesharing demand and supply in Chicago City. Additionally, adding external factors such as extreme weather conditions, special events and public transportation schedules in the prediction models could enhance the accuracy of forecasting.

REFERENCES

Chebance, Z., Markov, I., Guglielmetti, R., & Laumanns, M. (2021). Performance Comparison of Supply–Demand Matching Policies for On-Demand Mobility Services. *Transportation Research Record*, *2675*(11), 1-12.

Chicago Data Portal. (2023). Retrieved March 20, 2023 from https://data.cityofchicago.org/

Chung, J. C. G. K. C., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv*, *1412.3555*.

Dey, B. K., Tirtha, S. D., & Eluru, N. (2021). Transport Networking Companies Demand and Flow Estimation in New York City. *Transportation Research Record*, *2675*(9), 139-153.

Faial, D., Bernardini, F., Meza, E. M., Miranda, L., & Viterbo, J. (2020). A methodology for taxi demand prediction using stream learning. 2020 International Conference on Systems, Signals and Image Processing (IWSSIP),

Gulcehre, C., Kyunghyun Cho, Razvan Pascanu, Yoshua Bengio. (2014). Learned-norm pooling for deep feedforward and recurrent neural networks. Machine Learning and Knowledge Discovery in Databases:: European Conference, Nancy, France.

Hao, Z., He, L., Hu, Z., & Jiang, J. (2020). Robust vehicle pre‐allocation with uncertain covariates. *Production and Operations Management*, *29*(4), 955-972.

Ke, J., Zheng, H., Yang, H., & Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging technologies*, *85*, 591-608.

Li, J., & Allan, V. H. (2020). Balancing Taxi Distribution in A City-Scale Dynamic Ridesharing Service: A Hybrid Solution Based on Demand Learning. 2020 IEEE International Smart Cities Conference (ISC2),

Li, T., Qi, G.-J., & Stern, R. (2022). Taxi Utilization Rate Maximization by Dynamic Demand Prediction: A Case Study in the City of Chicago. *Transportation Research Record*, *2676*(4), 367-379.

M. Schuster and K. K. Paliwal. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673-2681.

OpenWeather. (2023). Retrieved April 3rd, 2023 from https://openweathermap.org/

Qian, X., Ukkusuri, S. V., Yang, C., & Yan, F. (2020). Short-term demand forecasting for on-demand mobility service. *IEEE Transactions on Intelligent Transportation Systems*.

Sepp Hochreiter and Jürgen Schmidhuber. (1997). Long Short-term Memory. *Neural Computation*(9(8)), 1735-1730.

TensorFlow. *TensorFlow time series forecasting*. Retrieved October 15,2022 from https://www.tensorflow.org/tutorials/structured_data/time_series

Wali, B., Santi, P., & Ratti, C. (2022). A joint demand modeling framework for ride-sourcing and dynamic ridesharing services: a geo-additive Markov random field based heterogeneous copula framework. *Transportation*, 1-37.

Wang, D., Cao, W., Li, J., & Ye, J. (2017). DeepSD: Supply-demand prediction for online car-hailing services using deep neural networks. 2017 IEEE 33rd international conference on data engineering (ICDE),

Zhang, C., Zhu, F., Lv, Y., Ye, P., & Wang, F.-Y. (2021). MLRNN: taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis. *IEEE Transactions on Intelligent Transportation Systems*.

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu., J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68-75.

DECISION SCIENCES INSTITUTE

Fraud Reduction in Humanitarian Supply Chain Using Blockchain Technologies

Gillian Doby
Independent Researcher
gd.frances@gmail.com

Yu Amy Xia
The College of William and Mary
Amy.xia@mason.wm.edu

ABSTRACT

This research investigates fraud in humanitarian non-governmental organizations (NGOs) operations and suggests blockchain technologies to reduce NGO frauds through case studies and managerial analysis. We identify the roots of NGO frauds through cases and argue that the current NGO anti-fraud systems, which build on pro-active job training and strict rules as well as reactive heavy auditing processes, though functional, lack effectiveness and efficiency. We propose the prevention, easy task first, network-connection (PEN) strategies. We argue that when human decisions are replaced with secured blockchain records and automatic-administrated smart contracts, NGO intrinsic ethics are safeguarded to achieve the core missions.

KEYWORDS: Humanitarian organizations, NGO operations, Fraud reduction, Ethical behavior, Blockchain technology, Smart contract

INTRODUCTION

Humanitarian organizations, such as United Nations organizations (UNOs) and non-governmental organizations (NGOs), in which dedicated employees carry out stated missions and receive ample contributions from donors worldwide, provide necessary aids to people who are suffering and struggling. Humanitarian organizations perform critical roles in the development of our society and the improvement of our communities. The impact of these organizations can be seen in many ways – one of which is the recognition they receive for their valuable contributions. For example, the UN World Food Programme (WFP) was awarded the Nobel Peace Prize in 2020 for its efforts to combat hunger and improve conditions for world peace (BBC News, 2020).

Unfortunately, humanitarian organizations are also very familiar with the impact of frauds in their supply chain operations. UNOs and NGOs are in the position to attract individuals driven to help others, but also have the potential to entice unethical individuals who wish to take advantage of those who are unable or too afraid to fight back. In addition, the emergency situations in which these organizations provide aid, including fallout from hurricanes and pandemics, provide further opportunity for exploitation if activities are not monitored (Charles, 2016 & FCC, 2020). Consequently, humanitarian organizations must deal with dramatic fallout of internal and external fraudulent activities or endure significant financial loss and damage to their reputations.

Challenges: The Urgency and Necessity to Reduce Fraud in Humanitarian Operations

According to the results of the 2022 Global Economic Crime and Fraud Survey conducted by PwC, the top four types of fraud are customer fraud, cybercrime, asset misappropriation, and bribery & corruption. When broken down by industry, 36% of government and public sector organizations experienced cybercrime, 33% experienced accounting and financial statement fraud, and 28% experienced customer fraud including bribery and corruption (PricewaterhouseCoopers, 2022).

Frauds in humanitarian organizations can harm their reputation and erode public trust, leading to decreased support from donors and other stakeholders, and causing significant financial loss as well as widespread reputation damage (BBC News, 2014; New York Times, 2021). This can have a significant impact on the ability of NGOs to carry out their missions and deliver essential services to communities in need. Additionally, frauds can divert critical resources away from their intended purposes, reducing the effectiveness of aid programs and undermining the efforts of NGOs to create positive change in the world. Every incident of fraud in humanitarian supply chains results in fewer beneficiaries receiving the aid necessary to survive and improve their quality of life. Therefore, it is important for NGOs to have strong internal controls, transparency, and accountability measures in place to prevent and detect fraud, and to respond quickly and effectively if it occurs. By maintaining the integrity of their operations, humanitarian organizations can ensure that their resources are used effectively and that their efforts have the maximum impact possible.

It can be very challenging to identify unethical and fraudulent activity during an emergency due to the rushed nature of the operational response, limited documentation, and risk of damage to the already weak transport and communication systems. In the recent pandemic, there were several concerns pertaining to the release of COVID-19 vaccines through NGO operations. There were concerns about vaccines being stolen and sold on the black market and about fake vaccines being sold as genuine (INTERPOL, 2020 & Hopkins, 2020).

Frauds can nourish unethical working environment and lead to organizational corruption. Examples of this from Global Fund's reporting include cases in their offices in Nigeria and Indonesia, which were found to have kick-back programs requiring employees to return portions of their salaries to subsidize the salaries of superiors between April 2017 and October 2018 (The Global Fund, 2020). In Haiti, with the support from colluding managers of Global Fund, a third-party supplier overstated the size of the warehouse facility they were supplying for the storage of bed netting, resulting in the Global Fund overpaying a grand total of USD $216,870 in non-compliance expenditures (The Global Fund, 2019).

The efforts to eliminate frauds in humanitarian operations have always been thorough. Many NGOs not only construct organizational training systems and build strict internal operation rules, but also carry out frequent routine and ad-hoc auditing processes to evaluate different contexts and recommend improvement plans to lower fraud risks. In one such audit, the Office of the Inspector General (OIG) for the Global Alliance for Vaccines and Immunizations (Gavi) found the Republic of Guinea to be high risk in their 2019 report and determined a range of steps to be taken to reduce the risk. Similarly, in a 2019 report for the World Food Programme (WFP) on their food procurement process, WFP's OIG found the risk to be medium to high and included an improvement plan in their findings to ensure that the risk was decreased. However, the global organizations must manage the challenge of improving their methods and reducing fraud in multiple countries with different cultures, laws, and business practices, where auditing may be difficult, and information may not be transparent. Given the enormous challenges and the tremendous high difficulty level, fighting fraud remains an ongoing effort in NGO operations.

Remedy: Blockchain and Well-Designed Smart Contacts can be a Solution

We believe that operation transparency and safeguarded intrinsic ethics can significantly lower the chance of fraud and even prevent frauds from happening. NGOs would greatly benefit from improved technologies 1) that are made easily available for humanitarian organizations and their partners; 2) that can effectively monitor NGO internal and external supply chain activities; 3) that can significantly enhance supply chain transparency and support automatic-administrated contracts; 4) that can systematically prevent frauds from occurring while increasing operational efficiency and building more ethical operating environment.

In this research, we argue blockchain technology can meet the above requirement of new technologies, and we thus propose a strategy to adopt blockchain technology in NGO operations to reduce fraud and ensure ethical behavior in the complicated global operation environment. In particular, we argue that the transparent information sharing through blockchains along with the related smart contracts, which automatically generate and process agreements without human intervention, can limit the chances of frauds from happening and increase the robustness of the system to catch fraud in NGO operations. NGOs can promote an ethical working environment while being cost efficient, as the majority of their funding still goes towards their mission to help beneficiaries.

Blockchain technology can be a solution to reduce frauds in NGOs because it addresses many areas of concern in current NGO operation systems. It allows easy access to information along all points of a supply chain, without the added cost or issues associated with duplication of data. Blockchain technology also allows continuous control over the access of different types of users to unencrypted information. Moreover, the structure of blockchain systems prevents bad actors from alternating data in the system, making it easier to identify fraudulent activity. In addition, smart contracts built on blockchains allow activities and transactions be automatically administrated with pre-designed written code when pre-agreed conditions are met during the process. Without human intervention, unethical behaviors in humanitarian operations cannot find opportunities to exploit the system.

Blockchain technology can bring many additional benefits. First, blockchain as an open ledger allows easy external verification of the NGOs processes. This further defends against fraud beyond alternative auditing methods and increases NGOs operation transparency. Second, blockchain allows encryption of sensitive information so it can be available for verification without security risks. Third, smart contracts could be used to execute tasks more easily and securely while speeding up supply chain transactions. Fourth, smart contracts can be specifically designed to assist audits and monitor processes to lower human effort and reduce human error. Fifth, contractors can be automatically selected, and their performances can be evaluated and recorded without human bias. Some organizations such as UNICEF have already begun exploring the potential benefits of blockchain technology in their operations (UNICEF, 2021). This interest in practice indicates that NGOs would be receptive to the implementation of a new system.

Methodology

In our research, we work on promoting blockchain technology in humanitarian operations to reduce frauds. Case studies are used to examine past instances of NGO frauds and managerial analysis are conducted to develop our proposed application strategies to implement blockchain technology in NGO operations. To start, we study the structure of NGO supply chains by examining the supply chains of large humanitarian aid organizations. We then look at the occurrences of fraud in NGOs by going through reports of fraud in NGOs to find common trends and build a case analysis. Next, we explore blockchain technology and its uses in fraud prevention and delve into the benefits and challenges of the use of blockchain technology for NGOs. Finally, we design and strategize the adoption of blockchain technology and smart contracts by matching areas of weakness with blockchain solutions.

LITERATURE REVIEW

This research is closely related with the following streams of literature: (1) supply chain fraud control, (2) blockchain technology and its application to fraud management, particularly in humanitarian operations fraud control.

Supply Chain Fraud Control

NGO supply chains face a range of issues related to fraud control, including monitoring sustainability standards (Perez-Aleman and Sandilands, 2008), enhancing transparency on compliance with standards (Chen at al., 2018; Shao et al., 2020), data analysis to improve humanitarian supply chains functionality (Prasad, et al., 2018), etc. With the knowledge from the related areas, researchers have attempted to identify and find solutions to supply chain frauds. Van Ruth et al. (2017) study food fraud prevention and identify the intentional nature of fraud. They explore a new food fraud vulnerability concept, which is based on the criminological routine activity theory and 31 detailed fraud vulnerability factors. In a separate paper, Van Ruth et al. (2018) provide a breakdown of fraud drivers and enablers into different categories, with a focus on vulnerable food supply chains. This study identified thirteen motivations and opportunities that can lead to fraud and found the most vulnerable food supply chain to be in the spice industry and the most vulnerable supply chain actor to be wholesalers or traders. Other researchers explore means of fraud identification, while focusing less on the motivations behind it. Kraus and Valverde (2014) use data analysis to find a new method of fraud detection. They determine that supply chain data obeys Benford’s law, which observes that in many real-life sets of numerical data, the leading digit is likely to be small. With this observation, with proper data storage, Kraus and Valverde (2014) build a tool that could detect supply chain fraud and ease the workload of fraud analysts. Malini and Pushpa (2017) study different fraud detection and analysis techniques for credit card fraud. They evaluate how machine learning and fuzzy logic along with algorithms such as KNN or outlier detection, can help identify and prevent fraudulent transactions. Ellis et al. (2015) take a more physical, test-based approach to fraud identification when they promote handheld “point and shoot” devices to bio-chemically test food at different points in a supply chain to detect any counterfeit or contaminated goods.

Blockchain Technology and Fraud Management

Blockchain technology with its decentralized and secured ledger has drawn attention in fraud prevention in recent years. Blockchain technology has been adapted to different uses around the world. The exact functionality and user structure can change based on the type of blockchain and its purpose, but there are a few characteristics that are intrinsic to blockchains found in all variations. The most distinctive of these characteristics is immutability. Dai et al. (2017) lay out a few benefits of blockchain in terms of adding security against fraud. The short paper also briefly discusses the risk of managerial control over the inputs into a blockchain from the computer science perspective. Immutability is the quality of blockchain to be unchangeable after a transaction has been completed and recorded in the blockchain (Hofmann, et al. 2017).

This quality comes from the structure of blockchain. Blockchain is composed of blocks of stored information; each block is stored in one of the different nodes or computers that form the network on which the blockchain is built. Each block captures any new transactions on the blockchain as well as information from previously recorded blocks in the system. As a result, information for each transaction is locked in by information shared between multiple blocks. This makes it exceedingly difficult for any actor to change historical data, as the information between blocks would no longer align if someone were to attempt to tamper with the stored data (Hofmann, et al. 2017). Hence, blockchains are an excellent tool for security and fraud prevention, as all recorded information can be verified.

The structure of blockchains is also what makes them transparent. As information is stored in blocks spread across the computer network, the ledger is shared amongst members of the blockchain. All users can verify data within the ledger and confirm its authenticity. The widespread access to the data in the blockchain allows for high levels of transparency with all users. However, though all users can see data stored in the blockchain, it is still possible to maintain the security of information as necessary through the use of encryption (Atlam, et al. 2018). In this way, though all users can verify the authenticity of data, only authorized users can see the data in its unencrypted format.

Finally, smart contracts are the finishing feature of blockchain that have been increasingly explored. Smart contracts are code that is overlayed across a blockchain. Etherium has become the most commonly used blockchain host for smart contracts, but their control over the market has led to increased costs for the customer organizations (Liebkind, 2019). They typically take the form of "if/then" executions, in which the code is executed if the required conditions are met (IBM, 2021). By having smart contracts execute tasks independently, transactions can be completed more quickly and without human bias.

The above literature built a solid base for our exploration. However, most of them either do not adequately resolve the issue or do not specifically consider targeting humanitarian NGO fraud. Our paper's research delves into examples of fraud prevention across humanitarian organizations and propose a blockchain adoption strategy to reduce NGO frauds.

NGO FRAUDS AND THE FITNESS OF BLOCKCHAIN IN FIGHTING AGIANST FRAUDS

To accomplish their central missions, NGOs usually operate in a widespread global supply chain. In this section, we first analyze frauds in humanitarian operations with various cases, which highlight the weak spots where frauds often happen in NGO supply chains. We find that procurement processes and logistics systems are particularly vulnerable to frauds. We then discuss an example to locate potential frauds in NGO operation processes. In addition, we study procurement strategies and match the functionalities and characteristics of blockchain with NGO's anti-fraud procurement strategies. We discuss why and how blockchain technology can effectively and efficiently execute the anti-fraud strategies of NGOs.

Fraud Cases

In the subsection, we investigate instances of humanitarian frauds to understand their causes and identify the spots where they occur. Notably, the NGOs we examine are usually transparent with fraudulent activities; their internal investigating groups are alerted to frauds and routine audits are scheduled. Table 1 below briefly describes a sample of the cases. The greater description and details of the frauds are listed in Appendix A of this paper. Notably, many of the identified frauds took place during the procurement process in the form of falsified expenditures, counting both instances of overcharging and complete fabrication of expenses.

Table 1: Examples of Frauds and the Spots of Fraud

Organization	Country	Start Date	End Date	Value	Type of Fraud	Area of Operation
Global Fund	Haiti	November 2016	August 2017	USD $216,870	Collusion, Non-compliance with bid requirements	Procurement
Oxfam	United Kingdom	February 2011	December 2011	£64,613	Falsified Receipts	Procurement
GAVI	Niger	Not stated	2011	USD $1,446,372	Falsified Receipts, Ineligible expenses	Procurement
Global Fund	Nigeria	April 2017	October 2018	USD $ 166,930	Salary Kickbacks	Staff
Global Fund	Indonesia	April 2017	October 2018	USD $35,310	Salary Kickbacks	Staff
Global Fund	Sierra Leone	July 2016	August 2018	USD $2,331,127	Collusion, Inflated Expenses, Product Substitution, Non-Delivery	Procurement
WFP	Total Fraud	January 2019	December 2019	USD $9,894,285	All Types	All

To further study the above cases, we take the fraud that took place with Global Fund in Sierra Leone in 2018 as an example. The National AIDS Secretariat (NAS), one of the recipients of the Global Fund, selected four sub-recipients under the grant to assist with the vocational training efforts: Women in Crisis Movement, Society for Women and AIDS in Sierra Leone, Rofutha Development Organization, and Kakua Hospice. The four sub-recipients were only allowed to make minor procurements under a specified value; all major procurements were to be carried out by NAS. However, it was uncovered that NAS allowed their sub-recipients to make major procurements against grant specifications.

The non-compliant major procurements carried out by sub-recipients included procurement of vocational training for beneficiaries and procurement of start-up kits for graduates of the vocational training. NAS sub-recipients colluded with vocational institutions to increase the price per student to use the entire grant total. The investigation of the fraud determined that comparable or better vocational institutions listed student tuition at USD $60 per year in 2016 to 2017, however the vocational institutions selected by the NAS sub-recipients charged USD $138 per student in 2016 and USD $489 per student in 2017. Similar fraudulent activities were found in their procurement of start-up kits for vocational training. The NAS sub-recipients pre-selected vendors, circumventing the required bidding process, and then colluded with suppliers to inflate the per kit cost. Fraud was also identified in separate procurements under the same grant. One of these was the procurement of 138 desktop computers by NAS, where evidence of bid-rigging, fraud, product substitution, and non-delivery were found, with USD $16,113 worth of computers unable to be located. Frauds were also found in the procurement of USD $121,768 worth of medical furniture and equipment for drop-in centers. Table 2 lists some of the details of the frauds.

Table 2: Fraud Committed by NAS under Global Fund Grants in Sierra Leone			
Fraud	Examples	Cause	How Blockchain Can Help
Overcharging	-Vocational training -Start-up kits		Transparency & Smart Contract
Acting outside authorization	-NAS sub-recipients making large procurements	No technical prevention/authorized user access to system/funds	User Authorization: Only allows authorized users
Non-delivery or unverified delivery of goods	-Start-up kits -Medical equipment/furniture -Computers	Lack of documentation	Block payment until delivery documentation is completed in system
Supplier substitution	-Lower grade processor -Reduced RAM -Computers with no operating software	Lack of inspection	Transparency & Smart Contract Block payment until passed inspection is registered in system
Non-compliance	-Medical equipment/furniture procured did not match required specifications -Number of start-up kits did not match number of training graduates	Lack of contract/procurement oversight	Smart contract Ensure contract details match requirements to proceed
Bid-rigging and collusion	-Vocational training -Start-up kits -Computers		Smart contract

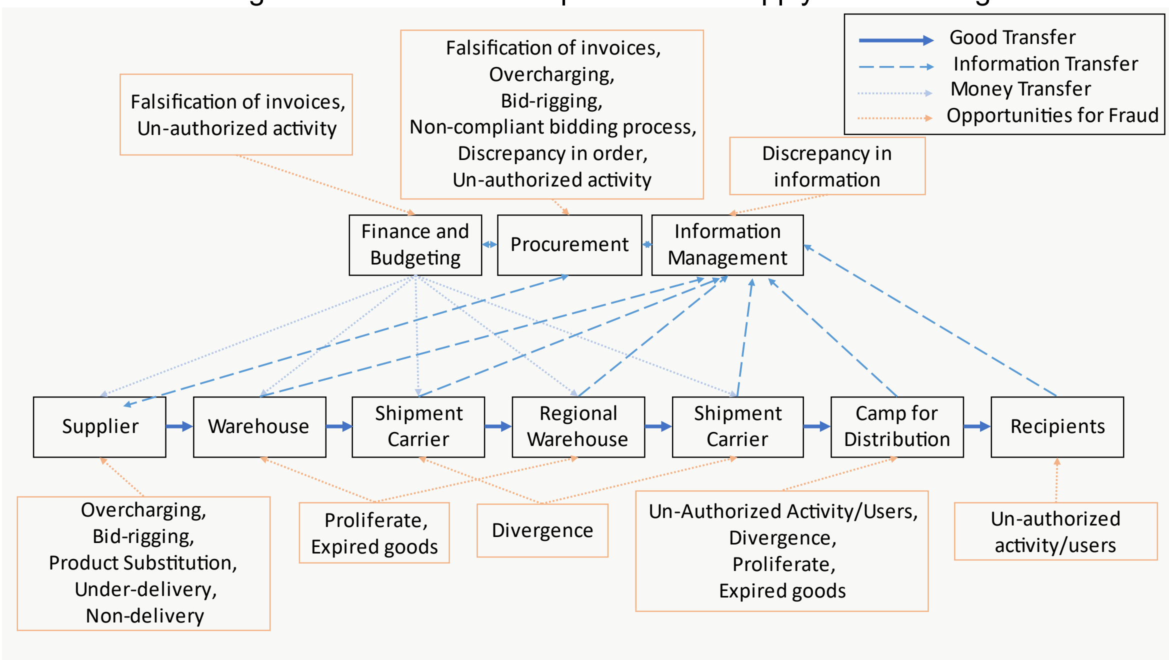
In the above case, NAS had lost control of the procurement procedures of the fund, their oversight allowed the fraud to take place among the subrecipients. In order to reduce frauds in procurement, they should have secured authorization for procurement and made sure that material flows in terms of item identification, quantity, and quality with the contract requirements match with information record. In addition, requirements of the procurements should be enforced throughout the system. Ideally, if an order does not satisfy all the requirements, it should be automatically identified and stopped.

The Spots Where Frauds Happen

In this subsection, we further analyze the NGO procurement strategies and identify what can be implemented through blockchain technology.

In Figure 1, three NGO operation functions, namely finance and budgeting, procurement, and information management are connected within the example supply chain structure and possible fraud areas are identified. For example, in negotiating supplier contracts, overcharging may happen within the contract while kickbacks go to individual arranging the agreement. Also, while money is transferred from budgeting to suppliers for payment, invoices may be falsified for personal benefit. In these possible fraud locations, if information can be kept in a secured and decentralized ledger that cannot be easily altered and is connected via signals of material flow (e.g., through Internet of Things), we can effectively prevent frauds from taking place.

Figure 1: Potential fraud spots in NGO supply chain management



Why Blockchain?

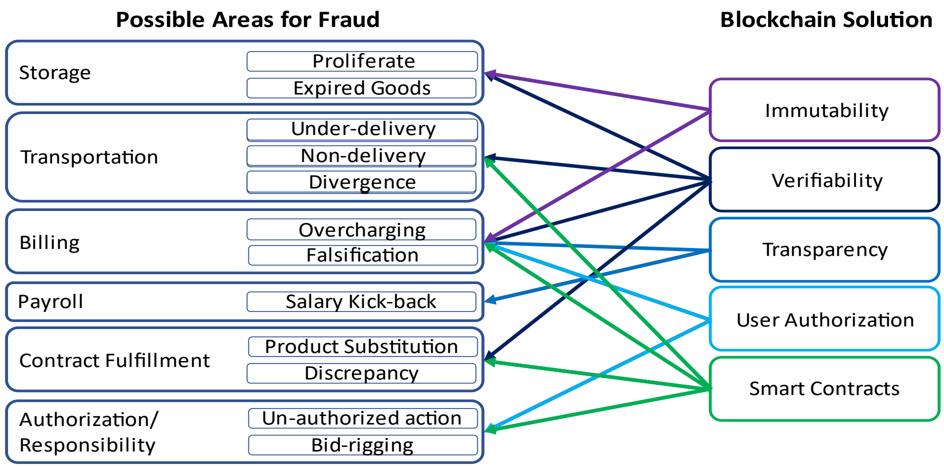
We argue blockchain technology can significantly help to prevent fraud in supply chain procurement processes, protect information, such as people's identities, while keeping a supply chain transparent and accountable.

Take a UNOs such as WFP as an example, their procurement principles include best value for money, competition, fairness and transparency, and interests of the organization (WFP Executive Board, 2019). In order to achieve their goals, they have strict policies in procurement. In term of authorization, the micro purchase orders (MPO) policy states that low value purchases, under $5000, can be managed by MPO focal points within relevant units on request (WFP OIG, 2021). Guidelines are also written to structure various areas of procurement including the number and evaluation of competing bids during a selection process (WFP OIG, 2021).

To go back to the fraud cases in Table 2, Column 3 illustrates the causes of the fraud, while Column 4 matches Blockchain characteristics with the functions that can reduce or prevent the causes from happening.

The combination of blockchain features can contribute to a simple and effective auditing process. Humanitarian organizations are frequently audited, both internally and externally by donors and other groups (UN Secretary-General, 2013), and therefore would benefit greatly from blockchain's capacity to make the process more efficient. The transparency of information improves access for auditors. Since the NGO's data would be encrypted (at least in part), auditors could be allowed access to audit all non-sensitive information within the system and could still verify that encrypted data had not been tampered with. The immutability and verifiability of information in a blockchain would improve the process of checking data during an audit. Finally, smart contracts could be built to present data in a relevant format for auditing or to perform specific checks for the auditors as needed, improving the speed and accuracy of audit results. Below is a graphic of blockchain solutions and the potential areas for fraud that can be addressed in humanitarian organizations:

Figure 2: Blockchain solutions match with the potential areas for frauds



There are other business solutions in current use that address many of the fraud prevention and identification issues that have not been fully detailed in this paper; these solutions include centralized databases, centralized ledgers, and cloud storage, among others (Lawton, 2021 and Joshi, 2021). These solutions work well in both the public and private sector; however, we posit that blockchain adds significant value in fraud solution for NGOs, who value trust with donors and the public. Blockchain allows NGOs to maintain the highest levels of transparency with external parties whom they rely on for donations and assistance while working on completing their missions. The public distrust of an organization can be drastically reduced when communication channels are open and information is made publicly available (Schnackenberg & Tomlinson, 2014). NGOs in particular face criticisms for not releasing information publicly on their projects. The immutability of blockchain makes it a valuable solution to any such criticisms of NGOs as any relevant information can be made available and verifiable to members of the public, preventing any potential concerns over false statistics.

When preventing fraud, it is important to consider the fraud triangle: incentives, opportunities, and attitudes (Johansson and Carey, 2016). The three sides of the triangle impact the occurrence of fraud in an organization by providing a deeper understanding of all components of a fraudulent act and thereby be better prepared to prevent those components and actions. The use of blockchain technology can be used to address at least two sides of the fraud triangle, incentives and opportunities. Incentives or motivations for potential perpetrators to commit fraudulent activity are a balance of benefits and consequences. If risk of being identified is higher, it reduces motivation to commit fraudulent activities.

Immutability is the distinguishing feature of blockchain, preventing data stored in the blockchain from being edited or hacked after the fact, significantly increasing system security. Immutability is possible within blockchain due to its decentralized verification and its stamp and hash process. When new data is added to a blockchain, the time is recorded as a stamp, the data is verified by multiple or majority nodes in the network (details depend on the protocol the blockchain chooses), and then the time stamp and the new data are hashed with all former information through an encrypted process to build a block (Gomi, 2021). The block is almost impossible to be altered, so is the information in the block. These features serve as the security foundation of a blockchain and can be built upon with other blockchain tools.

Transparency is another feature that makes blockchain ideal for NGOs. All the data in the blockchain can be public facing and publicly verifiable with sensitive data encrypted as needed. Transparency allows NGO for increased public trust (Jung, 2019). It also serves as an incentive for NGOs and the organizations they are working with to operate in a trustworthy manner, with the knowledge that all activities will be visible. In all the instances of overcharging seen in the case analysis, notably those in which the charged prices did not align with average costs in the market or inflation, a system with public transparency would have made the fraud much more challenging to hide and led to an earlier identification of the problems.

Authority and authentication are other essential tools within blockchain that can prevent fraudulent activity within a supply chain. An example of fraud preventable with authorization can be seen in our case analysis – the fraud was committed by the sub-recipients, who handled procurements without the proper authority to complete transactions of that size. With blockchain, users can be assigned authorizations according to their role in the supply chain (Mohsin, et al, 2018). The activities in the Global Fund case could have been prevented as the sub-recipients, who were the primary actors in all fraudulent actions in the case, would not have been able to complete the large procurements, regardless of whether fraudulent intensions had been present.

Since blockchain is effective security against cybercrime and the smart contract structures discussed in this paper are effective prevention against many forms of accounting fraud and corruption, blockchain could significantly reduce the frequency with which fraud occurs in the world of humanitarian supply chain. Delving the specifics of exactly how much of each type of fraud could be prevented with blockchain and well-structured smart contracts is an area of interest for future research.

The benefit and cost analysis of adopting blockchain in reducing humanitarian fraud is always difficult to measure. Once it can be projected how much fraud will be prevented, the full benefits of blockchain can then be compared to the costs of implementation and maintenance of blockchain in NGOs. This cost will vary between organizations depending on size, resources, and technology currently available. Our first suggestion to reduce the cost of implementation is to choose to build a permissioned blockchain. Permissioned blockchains can be scaled to an appropriate size for each organization and are typically more energy efficient than permissionless blockchains, helping adjust costs to organizational needs (Oracle Developer Resource Center, 2022). Next, we suggest looking at competitors to the top blockchain server systems to compare costs. By examining the blockchain market thoroughly before entering, it will be easier for NGOs to identify third party organizations that can host their blockchain system for less than the front running organizations.

PEN STRATEGY TO INCLUDE BLOCKCHAIN IN NGO OPERATIONS

There are many challenges to the implementation of blockchain in an organization – these include but are not limited to building a compatible system, gaining acceptance and participation from users, and training users in the new system. In this section, we propose the *Prevention, Easy task first, and Network-connection* (PEN) strategy for NGOs interested in implementing blockchain across their supply chains. The key to this strategy is to take a cautious and step-by-step approach to implementation, focusing on successful connection of all areas and all signals within the supply chain. In addition, we introduce smart contracts on top of blockchain technology to further facilitate humanitarian operations and reduce chances of fraud.

Prevention.

It is our recommendation that when taking a measured approach to the implementation of blockchain, the integration process should first and foremost focus on prevention. A focus on prevention means that items to be integrated into the blockchain are prioritized by the need for *authentication* to stop fraudulent activities from moving forward. For example, identifying personal information of beneficiaries, including biological identifications such as fingerprints or retinal scans, can be incorporated in the blockchain record and used for goods and cash distribution, reducing the potential for fake recipients to divert welfare by securing identification authentication in the system. The risk of benefit cards or vouchers being stolen or forged will largely be diminished. Biological identifications can also be imbedded in NGO human resource systems and pass through the blockchain to manage payment for overseas employees and contractors effectively and accurately. Conventional systems may also incorporate these biological identifications; however, biological identification information will be much better protected with the hash processes secure the information blocks in the blockchain system. The decentralized records in blockchain also lower the chance for the information being fabricated or altered as compared to a centralized system.

A prevention-centric approach should also focus on adding *automatic executions* of standard processes and pre-agreed transactions. It is important to reduce human intervention in basic functions in humanitarian supply chain operations as a fraud prevention measure. In addition, *condition-triggered alerts* can also be imbedded in blockchain NGO systems. These alerts can send warning signals to proper internal auditors or investigators and suspicious behavior can be flagged and suspended to avoid fraud. All prevention measures can be seamlessly incorporated through smart contracts or signal connections on top of secured blockchain records, which are almost impossible to alter.

Blockchain empowered supply chain financing is another fraud prevention measure. Inclusion of supply chain financing facilitates more efficient supply chain transactions and allows more reliable and prompt payment for NGO suppliers or transporters. Since blockchain can efficiently track and trace shipments and items when connected with network signals (IoT), with a decentralized and trustworthy blockchain record, financial institutions can process transaction-based loans much faster with less chance of fraud.

Easy Task First.

Instead of having an overall and systematic approach upfront, we propose to start with easy tasks and gradually improve and implement the blockchain system in NGO operations. We believe the overhaul of the current system is unnecessary, expensive, and problematic. Instead, a step-by-step approach is encouraged. Processes need to be organized based on complexity level and accessibility level. The implementation of the blockchain should start with the least complicated and most accessible processes first and move up to more complicated processes. In NGO practice, blockchain implementation can *start with front desk operation processes* that connect NGOs with suppliers, contractors and distributors, and receivers. These front-desk processes usually include data input and delivery recordings, are relatively simple and easy to access, and are critical in keeping the accurate and trustworthy records in the blockchain system. Also, these processes are usually already connected with the existing supply chain system. Adopting blockchain at these activities first can efficiently build the blockchain system through the supply chain and add security and trust through the system. By completing front desk operation processes, the blockchain system secures a few segments of the supply chain that are linked through the front desk, creating initial benefits and utility in the system. This order of inclusion also increases the level of security across the system early in the implementation process and allows for early successes to be communicated to management and donors. Once the optimal opportunities have been acted on, the more challenging components can be added.

Humanitarian blockchain implementation can be further *enhanced by building a basic database for verification* once the input data are verified at the front desk. The database can be based on existing structures and built with an understanding of which items need to be verified or accessed most often and how the system must be structured to contain all the essential information. The goal is to create a system that adds value without creating the need to reteach all concepts to the personnel, suppliers, and beneficiaries using it. Migrating to a blockchain based system can be used as an opportunity to clean up the existing structure, but the primary goal is to contain the necessary information and functions in a system that prevents alteration. We should *work on the existing design of the supply chain* system. More difficult or complicated tasks that incorporate and modify multiply steps involving multiple parties will follow.

Finally, we suggest *a backward approach* to integrate processes in the newly built blockchain system. Specifically, the integration should begin at the supply chain end-nodes at the receiver side and progress towards the beginning of the supplier side. This is due to the large quantities of data typically found at the end of a supply chain, including large numbers of distributors in the last mile of delivery and personal information from beneficiaries. As mentioned previously, an example of a good blockchain opportunity at the end of a supply chain would be the retinal scans utilized by organizations such as WFP to identify end recipients picking up supplies. These scans are identifiers at the very last step of a supply chain that need to be kept secure both for recipient privacy and to ensure designated recipients are receiving the supplies (WFP, 2016). Placing this information in an encrypted blockchain would allow for the safe keeping of the information with easy access for the end distributors that are performing the scans.

Network-Connection.

It is also imperative that there should be a strong *Internet of Things* (IoT) connection throughout the blockchain. Having a good connection with IoT will ensure accurate input of information and proper communication throughout the blockchain, it will also support fraud prevention efforts in verifying information and preventing fraud opportunities in the blockchain. A blocked connection would make quick and easy access of all parties impossible. To ensure a smooth and strong IoT connection, an immediate *signal verification* at the connecting point should be implemented, in addition to regular *document checks* throughout the blockchain, to confirm that the IoT connections are functional and effective. To safeguard the ability to perform checks that monitor connection and data quality through the system, it is essential that all activity in the system be associated with a *digital signature* for processing and completion. The digital signature should keep track of goods, services, equipment, and individuals in the system to keep records of associated activities. All records associated with confirming the completion or status check of an activity should carry a specific, verifiable signature. Verifiable signatures include pictures of final delivery, barcode or RFID scans at different warehouses along travel route, and inspection records.

Network-connection should not be limited to functional parties such as product suppliers, transportation partners, and supply chain material flow activities such as loading, shipping, storing, and distribution. Stakeholders such as donors, beneficiaries, and hosting governments should also be connected to the blockchain to access and verify blockchain information and verify and facilitate monetary flow and information feedback from donors, financial service institutions, and NGO internal processes.

Although a complicated network of parties should be connected within the blockchain, the authority level of each connecting party should be clearly defined and pre-approved. For example, donors can be given access to financial data to verify activity expenses from their donation while beneficiaries should be able to verify their identities, usually with biological signatures, and the reception of the support.

SMART CONTRACTS

Smart contracts are codes embedded in blockchain and can automatically execute an action or a transaction when specified conditions are verified through the connected network signals (IMB, 2021). Using smart contracts to control and flag behavior in the system allows for thorough and fully customized security.

Smart contracts are one of the most flexible and value adding tools in blockchain that gives an organization the most control over stopping fraudulent activity and tailoring a blockchain to their needs. For example, when goods are delivered, the payment for the goods can be locked until they have been inspected. Then, after an authorized user has recorded in the blockchain that the goods delivered are the correct items, with all required components and in the correct quantity, the smart contract code could automatically release the payment. The code can be specialized to a variety of different uses and needs, allowing for processes to be better secured against frauds with the added benefit of automating certain processes for system users. Though auditing would still be necessary to ensure that the system catches all fraudulent activity, it would be possible to use blockchain smart contracts to create alerts, triggered by suspicious activity, so that employees and auditors can know where to start looking when the system cannot prevent it directly.

User Identity Verification

The first, though perhaps most frequent, use of smart contracts will be verifying the identities of all users, items, and locations in the system, at every instance of use. For the system users, including donors, employees, suppliers, distributors, and transporters, the use of their corresponding identification (ID) number will trigger a smart contract in the blockchain. The smart contract will match their identification number to their assigned role in the system. If the user is permitted the action they are attempting, the smart contract will allow them to proceed, otherwise their actions are blocked, and a flag is created in the system. This will ensure that only permitted users with assigned authority are performing actions in the system as one method of avoiding frauds. Using the Sierra Leone case example, the sub-contractors would have been stopped from completing a procurement outside their permitted price range by the smart contract.

Transportation Verification

Generating and verifying transaction ID numbers will also be useful when tracking items in a supply chain. An ID number would be generated to each storage facility and product; as a precaution, ability to generate product IDs should be limited to specific user roles. When a transporter scans an item, it should register the transporter ID, the warehouse ID, and the product ID. The blockchain can then keep an easily accessible record of who has handled the product and where the product has been. Using this information, a smart contract will create a flag if a product is handled by an unexpected transporter or arrives at a warehouse outside of its planned path, bringing the potential error or fraud to the attention of the organization.

Expiration Dates

Keeping track of expiration dates associated with each product ensures both that goods are moved and used before their expiration dates and also ensures that there is no fraudulent substitution of expired goods during procurement or transport. Using a smart contract to indicate a warning when goods are near their expiry date assists with the issue of using goods prior to their expiration. In order to ensure goods have not been exchanged with expired goods, it will be necessary to pair these smart contracts with inspections. Any discrepancies between listed expiry dates and those found in inspections would be flagged by smart contracts for review by the organization.

Payment

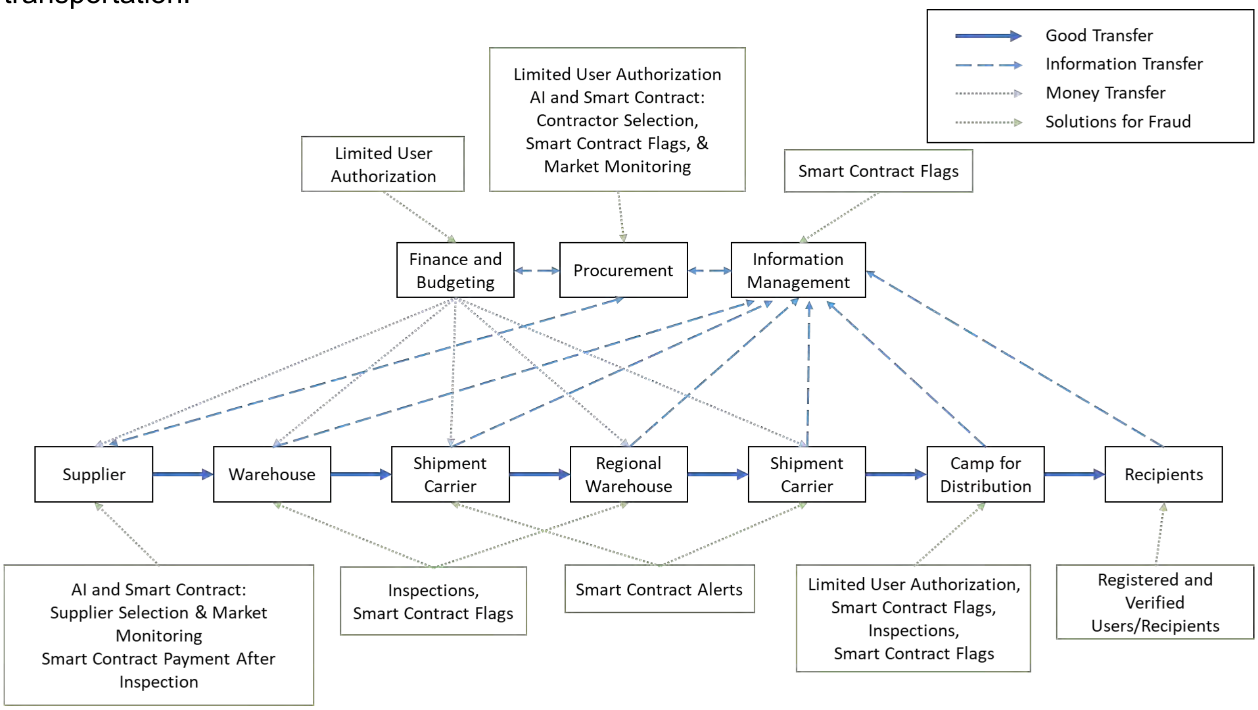
Payments should be tied to smart contracts. When payment information is entered into the blockchain, it will include the cost/price, contractor account information, and requirements (material flow IoT signals) for payment. For example, the requirement for transporters would be on time delivery of items in good condition. As soon as the specified signals of successful delivery are received, correct payment should be processed through a smart contract to guarantee the payment to the suppliers. The smart contract executed payments can add security and trust in the blockchain-enhanced humanitarian supply chain, and thus facilitate the NGO transactions with lower cost and lower chance of fraud.

Supplier and Transporter Selection

Smart contracts, potentially through artificial intelligence algorithms, should also be put in place to prevent bid-rigging in the contractor selection process. The blockchain bidding system should be open externally for bid submission by potential transporters and suppliers. The smart contract can perform the first filtering process, eliminating bids that do not meet contract needs or are well above average market cost through its automatic and pre-designed calculation. The smart contract can even select the winning bid that have the capacity to complete the contract and have a lower probability of collusion based on its algorithm. A smart contract can also stop a procurement process to advance or flag it if it doesn’t meet required protocol. Though human inputs would be required for certain features of the bidding process, overall, it would greatly reduce the risk of bid-rigging, human bias, and fraud in the procurement process.

Inspections

Since the blockchain can only monitor the information recorded in the system, the importance of continued inspections for goods and services should be emphasized. These inspections need to take place when goods and services are received and also at set or random intervals throughout storage and transport. This is to ensure the received goods and services are correct, meet all contract requirements, and monitor any pilfering occurring in warehouses and during transportation.



CHALLENGES AND REMEDIES

Implementing and maintaining blockchain in humanitarian operations can require significant investments in terms of infrastructure, expertise, and ongoing operational costs. Other issues such as the complexity of integration can be challenging and time-consuming (Saberi et al., 2018). Blockchain technology itself is also restrained by its scalability and interoperability limits (Oracle Developer Resource Center, 2022). From management perspective, blockchain is a relatively new technology, and there is a lack of understanding and awareness among NGO staff about its potential benefits and use cases (Saberi et al., 2018).

These challenges are largely tied to the current system structure, desired scale and features, and operating locations. Organizations can limit complexity by planning the new system to best fit their operations and blockchain capabilities in ways that limit redesign for the current system and align with staff current functions and entry methods. NGOs should carefully evaluate the investment they want to make in their systems and map the requirements that would be needed to reach that goal to determine if their network, system, and security can support and benefit from the integration of blockchain technology.

Uncertainty and inconsistent regulations in a changing global economic and political environment as well as the frequent changes to technology and security standards (Nickerson, 2019; Button and Cross, 2017) can also create hesitation and reluctance for NGOs to adopt blockchain in their global operations. Adapting to the current global status of economics, politics, and technology is an ongoing challenge that requires constant monitoring and updates. Investing in a particular technology when new changes are introduced with increasing frequency is an organizational risk. Though blockchain does address concerns regarding security and transparency, it would benefit NGOs to examine the fit of the technology with their internal and external functions and needs. By thoroughly mapping the implementation for their organization, they can ensure the value added is sufficient to merit moving forward.

CONCLUSION

Based on the innate protective characteristics of blockchain and the tailoring availability through smart contracts, we recommend that humanitarian organizations consider using blockchain technology to ensure ethical behavior and defend their supply chain operations against fraudulent activities. We recognize there are currently constraints created by high costs and technology development hurdles, which may discourage or prevent organizations from taking advantage of this option immediately.

In this paper, we propose a *Prevention, Easy task first, and Network-connection* (PEN) strategy along with smart contract in blockchain adoption to overcome the above hurdles. We believe several positive outcomes can happen and create a bright future for NGOs to fight with fraud. Specifically, blockchain's transparent and immutable nature can provide greater visibility into financial transactions and resource allocation within NGOs. Implementing blockchain for fraud reduction can build trust and confidence among donors and beneficiaries and amid material suppliers and logistic service providers. The transparent and auditable nature of blockchain can assure donors that their contributions are being used for their intended purposes, leading to increased support and sustained funding. The smart contract associated with blockchain can guarantee payment for suppliers and service providers, and thus build trust, lower cost and facilitate transactions in NGO supply chain operations. Blockchain and the associated smart contracts can streamline and automate processes related to financial management as well as procurement within NGOs and in their supply chains. This can reduce administrative overhead, eliminate intermediaries, and ensure more efficient and accountable resource utilization. In addition, improved accountability, secured data sharing, more efficient measurement and reporting can also be the results of blockchain adoption and lead to humanitarian operations with reduced frauds.

DISCLAIMER:

The views expressed herein are those of the author(s) and do not necessarily reflect the views of the World Food Programme.

REFERENCES:

Atlam, H.F., Alenezi, A., Alassafi, M.O. and Wills, G., 2018. Blockchain with internet of things: Benefits, challenges, and future directions. International Journal of Intelligent Systems and Applications, 10(6), pp.40-48. Retrieved from: https://eprints.soton.ac.uk/421529/1/Published_Version.pdf

BBC News, 2014. Oxfam ex-fraud chief jailed for scamming charity. Available at: https://www.bbc.com/news/uk-england-oxfordshire-27588466.

BBC News, 2020. Nobel Peace Prize: UN World Food Programme wins for efforts to combat hunger. BBC News. Available at: https://www.bbc.com/news/world-54476569.

Button, M. and Cross, C. 2017. Chapter 4: Technology and fraud – The 'fraudogenic' consequences of the Internet revolution. Edited by McGuire, M.R. and Holt, T.J. The Routledge Handbook of Technology, Crime and Justice pp. 78-92.

Chen, S., Zhang, Q. and Zhou, Y.P., 2019. Impact of supply chain transparency on sustainability under NGO scrutiny. *Production and Operations Management*, *28*(12), pp.3002-3022.

Dai, J., Wang, Y. and Vasarhelyi, M.A., 2017. Blockchain: an emerging solution for fraud prevention. *The CPA Journal*, *87*(6), pp.12-14.

Ellis, D.I., Muhamadali, H., Haughey, S.A., Elliott, C.T. and Goodacre, R., 2015. Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis–moving out of the laboratory and into the food supply chain. *Analytical Methods*, *7*(22), pp.9401-9414.

FCC (Federal Communications Commission USA), 2020. After Storms, Watch Out for Scams. Federal Communications Commission. Available at: https://www.fcc.gov/consumers/guides/after-storms-watch-out-scams [Accessed June 2021].

Gavi, The Vaccine Alliance, 2019. Republic of Guinea's Programme Audit Report June 2019. https://www.gavi.org/sites/default/files/audit/pa/Programme%20audit%20report%20-%20Guinea-Bissau%2C%20March%202019.pdf

Gomi, K. (2021). Blockchain: Time-Stamping, Smart Contracts And Time-Stomping. Retrieved from https://www.forbes.com/sites/forbestechcouncil/2021/04/22/blockchain-time-stamping-smart-contracts-and-time-stomping/.

Hofmann, F., Wurster, S., Ron, E. and Böhmecke-Schwafert, M., 2017, November. The immutability concept of blockchains and benefits of early standardization. In 2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K) (pp. 1-8). IEEE. Available at: https://www.inno.tu-berlin.de/fileadmin/a38335100/PDF_Dateien/Publikationen/T-PROC-KALEI-2017-PDF-E-Hofmann.pdf

Hopkins, J.S., 2020. Covid-19 Vaccines to Be Stored Secretly Under Tight Security. The Wall Street Journal. Retrieved from: https://www.wsj.com/articles/covid-19-vaccines-to-be-stored-secretly-under-tight-security-11603278002.

IBM, 2021. What are smart contracts on blockchain? IBM. Available at: https://www.ibm.com/topics/smart-contracts [Accessed July 24, 2021].

INTERPOL, 2020. INTERPOL warns of organized crime threat to COVID-19 vaccines. INTERPOL. Available at: https://www.interpol.int/en/News-and-Events/News/2020/INTERPOL-warns-of-organized-crime-threat-to-COVID-19-vaccines [Accessed June 2021].

Johansson, E. and Carey, P., 2016. Detecting fraud: The role of the anonymous reporting channel. Journal of business ethics, 139(2), pp.391-409.

Jung, T. (2020). How transparency through blockchain helps the cybersecurity community. Retrieved from https://www.ibm.com/blog/how-transparency-through-blockchain-helps-the-cybersecurity-community/.

Kraus, C. and Valverde, R., 2014. A data warehouse design for the detection of fraud in the supply chain by using the Benford's law. *American Journal of Applied Sciences*, *11*(9), pp.1507-1518.

Liebkind, J. 2019. 4 Blockchain Contenders in Competition with Ethereum. Investopedia. Available at: https://www.investopedia.com/news/4-blockchain-contenders-competition-ethereum/ [Accessed July 13, 2021].

Malini, N. and Pushpa, M. 2017. Analysis on credit card fraud identification techniques based on KNN and outlier detection. In *2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)* (pp. 255-258). IEEE.

Mohsin, A. H., Zaidan, A. A., Zaidan, B. B., Albahri, O. S., Albahri, A. S., Alsalem, M. A., Mohammed, K. I. 2018. Blockchain authentication of network applications: Taxonomy, classification, capabilities, open challenges, motivations, recommendations and future directions. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0920548918303477.

Nikerson, M.A. 2019. Fraud in a World of Advanced Technologies. Retrieved from: https://www.cpajournal.com/2019/07/01/fraud-in-a-world-of-advanced-technologies/.

Oracle Developer Resource Center, 2022. Permissioned blockchain. Retrieved from https://developer.oracle.com/learn/technical-articles/permissioned-blockchain.

Perez-Aleman, P. and Sandilands, M., 2008. Building value at the top and the bottom of the global supply chain: MNC-NGO partnerships. California management review, 51(1), pp.24-49.

Prasad, S., Zakaria, R. and Altay, N., 2018. Big data in humanitarian supply chain networks: a resource dependence perspective. Annals of Operations Research, 270(1-2), pp.383-413.

PricewaterhouseCoopers, 2022. PwC's Global Economic Crime and Fraud Survey 2022. https://www.pwc.com/gx/en/forensics/gecsm-2022/PwC-Global-Economic-Crime-and-Fraud-Survey-2022.pdf

Saberi, S., Kouhizadeh, M., Sarkis, J. and Shen, L., 2018. Blockchain technology and its relationships to sustainable supply chain management. Retrieved from: https://www.tandfonline.com/doi/full/10.1080/00207543.2018.1533261?src=recsys.

Schnackenberg, A.K. and Tomlinson, E.C., 2014. Organizational Transparency: A New Perspective on Managing Trust in Organization-Stakeholder Relationships.

Shao, L., Ryan, J.K. and Sun, D., 2020. Responsible Sourcing under Asymmetric Information: Price Signaling versus Supplier Disclosure. *Decision Sciences*, *51*(5), pp.1082-1109.

The Global Fund, 2019. The Global Fund: Office of the Inspector General. Retrieved from: https://www.theglobalfund.org/media/8502/oig_gf-oig-19-012_report_en.pdf?u=637319002165600000.

The Global Fund, 2020. The Global Fund: Office of the Inspector General. Investigation Report: Salary fraud and abuse affecting Global Fund grants. Retrieved from: https://www.theglobalfund.org/media/9386/oig_gf-oig-20-004_report_en.pdf?u=637319003289000000.

USAID, 2019. Southern Africa - Tropical Cyclone Idai - Fact Sheet #10 FY2019: May 2, 2019. Retrieved from https://www.usaid.gov/cyclone-idai/fy19/fs10.

United Nations Secretary-General, 2013. ST/SGB/2013/4 - E - ST/SGB/2013/4 -Desktop. Financial Regulations and Rules of the United Nations: Secretary-General's Bulletin. Available at: https://undocs.org/en/ST/SGB/2013/4.

Van Ruth, S.M., Huisman, W. and Luning, P.A., 2017. Food fraud vulnerability and its key factors. *Trends in Food Science & Technology*, *67*, pp.70-75.

Van Ruth, S.M., Luning, P.A., Silvis, I.C., Yang, Y. and Huisman, W., 2018. Differences in fraud vulnerability in various food supply chains and their tiers. *Food Control*, *84*, pp.375-381.

WFP Executive Board, 2019. Local and regional food procurement policy. Retrieved from https://docs.wfp.org/api/documents/WFP-0000108552/download/.

WFP OIG, 2020. Office of the Inspector General Annual Report 2019. Retrieved from: https://docs.wfp.org/api/documents/WFP-0000118173/download/?_ga=2.246852612.1240379358.1600080679-1171849310.1599477592.

WFP OIG. 2021. Internal Audit of Goods and Services Procurement in WFP. Retrieved from: https://docs.wfp.org/api/documents/WFP-0000128469/download/.

APPENDIX A

Descriptive Table of Frauds in Humanitarian Organizations

Organization	Country	Start Date	End Date	Value	Description	Type of Fraud	Area of Operation
Global Fund	Haiti	November 2016	August 2017	USD $216,870.00	"Senior Supply Chain Manager of the Global Fund's lead implementer in Haiti, Organization Haïtienne de Marketing Social pour la Santé (OHMaSS), colluded with the Director of a third party supplier, JIT Logistics and Transport Company, to obtain a contract for the provision of a warehouse..." - Warehouse was smaller than stated - Manager and Director exchanged multiple bid drafts throughout bidding process - Resulted in overpaying per square meter and additional rent costs when additional space was needed	Collusion, Non-compliance with bid requirements	Procurement

Oxfam		February 2011	December 2011	£64,612.58	Edward Mckenzie-Green, "former head of Oxfam's counter-fraud unit has been jailed for defrauding the charity out of nearly £65,000." - 2 year, 5 month prision sentence - pled guilty - made payments for false invoices to fictitious firms	Falsified Receipts	
GAVI	Niger		2011	USD $ 1,446,372.00	GAVI conducted an assessment of Niger's funds management in 2011 that indicated fraud. GAVI then, with the support of the Niger government and ministry of health, conduced an investigation that revealed fraudulent activity and insufficiently documented expenditures, ineligible expenses, and unjustified disemburtsments	Falsified Receipts, Ineligible expenses	
Global Fund	Nigeria	April 2017	October 2018	USD $ 166,930.00	Salary kickback scheme (staff had to pay back a portion of their salaries to the organization once a month at the instruction of senior management)	Salary Kickbacks	Staff
Global Fund	Indonesia	April 2017	October 2018	USD $35,310.00	Salary subsidy scheme, staff required to return part of their salary, most of which went to the Executive Director	Salary Kickbacks	Staff
Global Fund	Sierra Leone	July 2016	August 2018	USD $ 2,331,127.00	Fraudlent purchases of goods and services under HIV/AIDS and TB/Malaria/Health System Strengthening grants - Non-compliant procurements under HIV/AIDS including bedrigging, overcharging, product substitution, and non delivery of goods - TB/Malaria/Health System Strengthening affected by fraudulent procurements and payments and overcharging for hotels and catering	Collusion, Inflated Expenses, Product Substitution, Non-Delivery	Procurement
WFP	Total Fraud	January 2019	December 2019	USD $9,894,285.00	Total losses from fraud found in 2019 from various sources, both internal and external, substantiated and estimated.	All Types	All

Sources for Fraud Cases:
The Global Fund
https://www.theglobalfund.org/media/8502/oig_gf-oig-19-012_report_en.pdf?u=637319002165600000
https://www.theglobalfund.org/media/9386/oig_gf-oig-20-004_report_en.pdf?u=637319003289000000
https://www.theglobalfund.org/media/9571/oig_gf-oig-20-011_report_en.pdf?u=637319003983100000
GAVI
https://www.gavi.org/sites/default/files/document/audit/GAVI_Alliance_Investigation_report_Niger_022012.pdf
World Food Programme
https://docs.wfp.org/api/documents/WFP-0000118173/download/?_ga=2.246852612.1240379358.1600080679-1171849310.1599477592
BBC News
https://www.bbc.com/news/uk-england-oxfordshire-27588466

DECISION SCIENCES INSTITUTE

From Waste to Wealth: A Future Paradigm for Plastic Management Using Blockchain Technology (BCT)

Nesreen El-Rayes
New Jersey Institute of Technology
E-mail: nde4@njit.edu

Aichih (Jasmine) Chang
New Jersey Institute of Technology
E-mail: jschang@njit.edu

Jim Shi
New Jersey Institute of Technology
E-mail: junmin.shi@njit.edu

ABSTRACT

This study addresses the escalating plastic crisis, integrating environmental, social, technological, and business perspectives, a viewpoint often overlooked in the literature. Utilizing Blockchain Technology (BCT), we propose a mathematical model to bolster recycling and enhance efficiency across the plastic value chain's upstream and downstream sides to close the loop. To this end, the model ensures systematic coordination and optimized interoperability, thus maximizing sustainability in Environmental, Social, and Governance (ESG) goals and Circular Economy (CE). In addition, it offers a novel perspective on leveraging technology to tackle global environmental challenges and elevate sustainable business practices.

KEYWORDS: Plastic Management (PM), Recycling, Blockchain Technology, ESG, Circular Economy

INTRODUCTION

Despite being introduced less than a hundred years ago, plastic has quickly become essential in nearly every industry because of its durability and affordability. The annual global plastic production increased from 50 Million Metric Tons (MMT) in 1976 to 390 MMT in 2021 (Statista, 2022). The surge in plastic production and consumption is triggering significant environmental repercussions. The three primary types of plastic waste in the U.S. are Low-Density Polyethylene (LDPE), Polypropylene (PP), and High-Density Polyethylene (HDPE), according to Statista, 2022. LDPE is commonly found in containers, packaging, food bags, and wrappings; PP is used in items like bottle caps, straws, and disposal cups and plates. HDPE is utilized in products such as bottles, toys, and bags. The primary plastics in Municipal Solid Waste (MSW) in the U.S. in 2018 are containers/packaging, durable goods, and nondurable goods (EPA, 2018). The top plastic waste items on beaches are plastic pieces, packaging/wrappers, bottles, straws/stirrers, and bags (El-Rayes et al., 2022). In comparison, the main waste items along the U.S. beaches are Cigarette butts, plastic bottles, and caps or lids (El-Rayes & Shi, 2022). Simultaneously, companies are actively embracing Industry 4.0 technologies (i.e., Artificial Intelligence (AI), Blockchain Technology (BCT), Internet of Things (IoT), and Cloud Computing (CC)) because of their various features that may promote sustainability. In particular, BCT is capturing the attention of its plethora of features, including anonymity, transparency, traceability, security, immutability, and irreversibility (See Chang et al., 2022, Chang et al., 2021, Kumar et al., 2020).

LITERATURE REVIEW

From 1990 to 2022, the term "Blockchain Technology" appeared 27,723 times in the Scopus Database, as noted by Chang et al., 2022. The countries leading in these publications are China, U.S., and India. BCT presents a promising avenue for sustainable supply chain Management (Rejeb et al., 2021). Although across various scholarly articles and technical reports, BCT, IoT, and AI have been cited as the leading three technologies with promising potential for leading sustainability (Chen et al., 2023; Walden, 2020; Kouhizadeh et al., 2020; Saberi et al., 2019; Sarkis et al., 2010). There is a dearth of research exploring the intersection of BCT and PM. A mere 43 publications address this subject, representing 0.004% of the total number of presentations related to BCT (El-Rayes et al., 2023), and the term BCT and CE coexisted in 163 publications between the years 1990-2022 (Chang et al., 2022). BCT applications for PM are still in the pilot phase, and many concepts need to be explored (Gong et al., 2022); while other scholars suggest that BCT for recycling is classified under different stages across the recycling process depending on the stage of the plastic life cycle and stakeholders involved (Ahmad et al., 2021). Several crucial factors determine the incorporation of BCT to foster a CE. These include the extent of automated production and the ratio of networked and controlled machinery throughout the supply chain (Bhubalan et al., 2022; Chaudhuri et al., 2022). At the same time, 80% of blockchain implementations are associated with modifications on the process level (Mougayar et al., 2015).

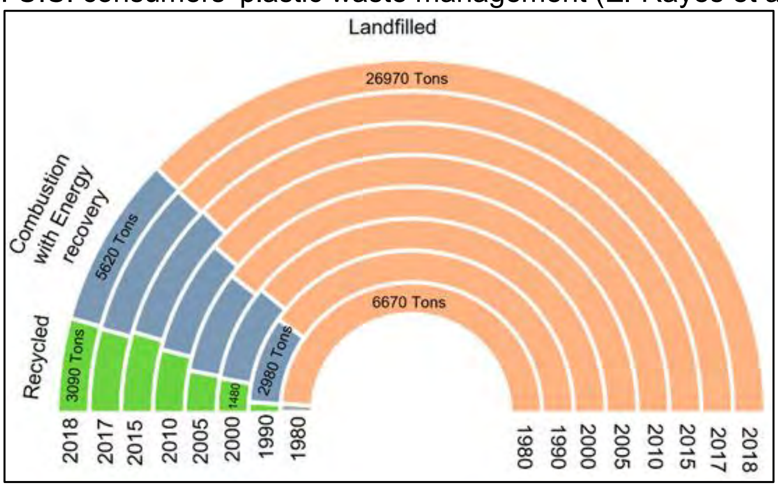
From late 2021 to 2022, the focus of publications on the Scopus database that we reviewed primarily focused on topics pertaining to marine debris and plastic debris removal. Hardesty et al. developed a model to identify the distribution and movement of minuscule plastic particles (0.02 to 5mm) in the ocean (Hardesty et al., 2017); Simulations help understand the distribution and trends of environmental issues (specifically marine litter) over more extended periods compared to the traditional surveys that rely on limited observations (Fossi et al., 2017). Monte Carlo simulations are incorporated to estimate the plastic pollution entering terrestrial and aquatic ecosystems (Lau et al., 2020).

While recycling seems like a crucial resolution to the plastic crisis, the recycling system in the U.S. is currently dysfunctional (El-Rayes et al., 2023) and faces numerous challenges. Chidepatil et al. propose an approach to handle plastic waste by integrating BCT with multi-sensors like visual sensors (VIS), near-infrared sensors (NIR), and far-infrared sensors (FIR) to extract the critical characteristics of plastic waste to classify plastics and determine the most suitable processing method or the fate for each plastic piece. Another approach suggests implementing BCT (leveraging crypto tokens) to collect, segregate, and sort plastic waste to enhance recycling efficiency (Khadke et al., 2021). Additionally, Bing et al. provide a decision support system to determine the most efficient combination or separation methods for recycling while considering the interest of stakeholders (i.e., municipalities or households) (Bing et al., 2014). The Waste Management Firm (WMF) is more likely to process a larger volume of waste when the market prices for processed wastes are high or when a more significant portion of waste has been preprocessed at the source (Chintapalli & Vakharia, 2023). Another approach employs Ethereum-based smart contracts to store data and provide meaningful information to stakeholders (Production & Quality control managers and existing & potential consumers) based on their needs (Nöbauer, 2022). Several functionalities may influence BCT's effective integration to address the plastic crisis. Some of which are the capability to sort and classify plastics according to polymer type, tracing waste from the point of creation to the point of management (i.e., tracking the historical usage of polymers), the degree of production automation, the proportion of machines digitally connected, the adaptability in modifying and adjusting processes with respect to the varying circumstances, and the seamless transition towards a digital platform (Bhubalan et al., 2022; Chaudhuri et al., 2022; Guyot Phung, 2019; Hristova, 2022; Katz, 2019; Nöbauer, 2022).

PROBLEM STATEMENT AND OBJECTIVES

The escalation of Global plastic production is a significant concern, evidenced by a 26% increase from 2010 to 2016 (Geyer et al., 2017). This growth trajectory suggests global plastics production will fluctuate between 500-600 MMT by 2025. Within the U.S., plastic waste generation is notably high per capita, with an average of 231 lbs. annually, surpassing all other nations (Forbes, 2020). Regrettably, the current waste management practices indicate that 77% of the plastic waste in the U.S. ends up in landfill, 16% undergoes incineration, and only 7% is subjected to recycling processes (Chaudhari et al., 2022). Figure 1 shows a detailed breakdown of the fate of consumer plastics in the U.S. from 1980 to 2018. That highlights the need for innovative solutions to increase the recycling rate and minimize the amount of plastic in landfill. On the other hand, there is a shortage of publications examining the inclusion of Industry 4.0 technologies to tackle the plastic crisis (El-Rayes et al., 2022).

Figure 1: U.S. consumers' plastic waste management (El-Rayes et al., 2022)



In that capacity, the objectives of this study are twofold: (1) Explore the research status of BCT in the realm of PM; (2) Examine the potential of incorporating BCT to enhance and optimize the efficiency of both upstream and downstream processes in PM.

RESEARCH QUESTIONS

(1) Could BCT help tackle the flaws in the recycling rate in the U.S.?
(2) What roadblocks, challenges, and actions are needed to promote BCT adoption and inclusion?
(3) What is the optimal value for the tokenization to maximize the plastic collection by the public?

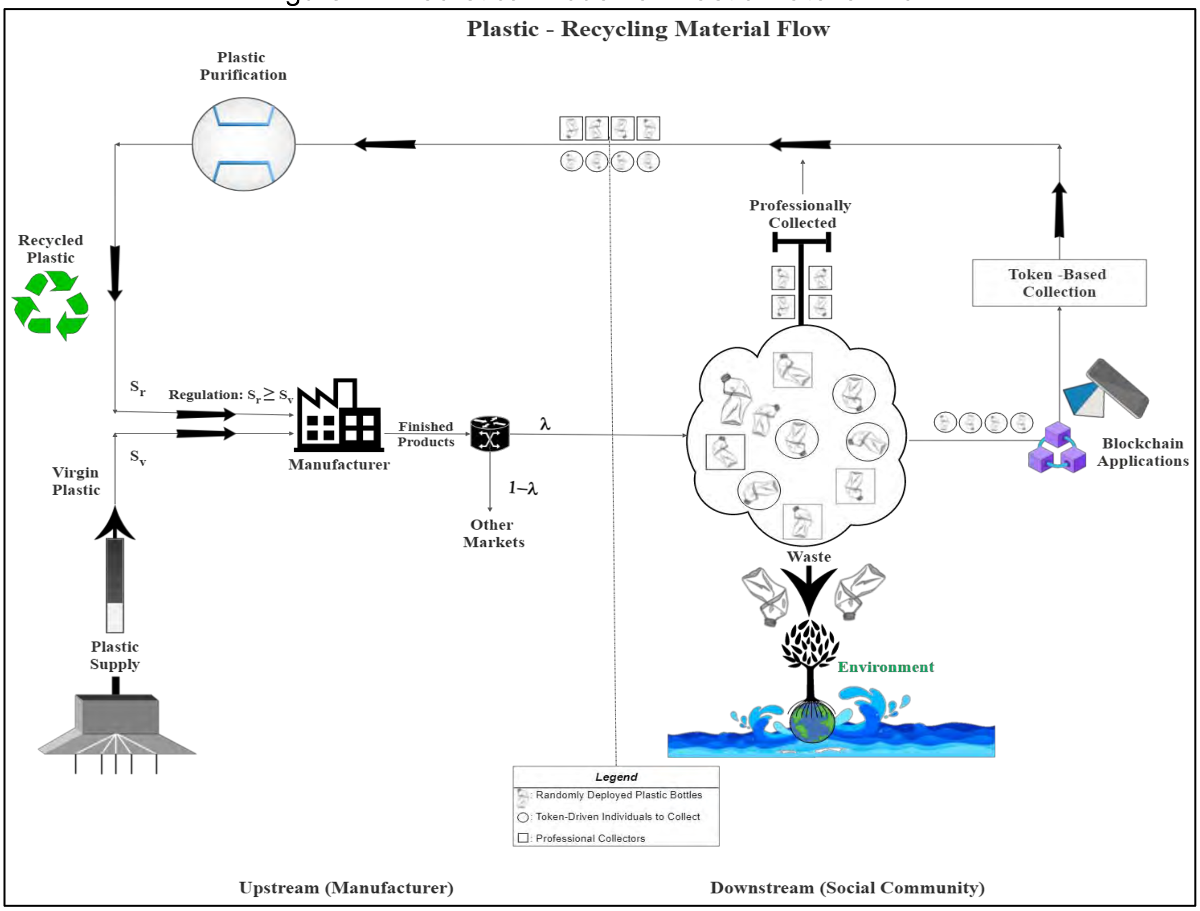
MODEL AND PRELIMINARY RESULTS

Data about plastic production, waste, and recycling was collected from the U.S. Environmental and Protection Agency (EPA) and the U.S. Department of State to comprehend the plastic crisis and construct an appropriate mathematical model. Mathematical modeling helps provide estimates, as sampling cannot fill gaps precisely when there is insufficient data (Daily, 2021). In that capacity, we create a dynamic programming model to glean valuable insights and develop actionable solutions. To close the loop, the downstream and upstream are examined as a whole ecosystem. In particular, we focus on the interoperability, system equilibrium, and sustainability of the ecosystem by which we show the benefits of Blockchain adoption. Finally, to address the urgent call on the CE, a BCT-enabled close loop system is proposed to enhance engagement and improve efficiency up and downstream of the value chain while maximizing the ESG goals.

Model Formulation

This model considers the plastic ecosystem within a loop between the downstream (individual picking up) and the upstream (plastic manufacturing). Figure 2 depicts the framework of the ecosystem. The decisions pertaining to the system are complex and intertwined. We focus on ESG objectives to capture the significant goal of different stakeholders.

Figure 2: Theoretical Model for Plastic Material Flow



Downstream Plastic Recycling

Considering a community (e.g., a public beach area) that randomly faces plastic waste disposal (e.g., plastic bottles). The community executes enforcement of its policy and campaign to restrict plastic disposals to some extent. Let $e \in [0, 1]$ denotes the enforcement scale. $D(e)$ denotes the disposed of plastic volume, which is random following any arbitrary pdf and $f_e(x)$ satisfying $f_e(0) = 0$ by convention.To incentivize individuals to pick up the disposed of plastic bottles, the community adopts a Blockchain-based token or cryptocurrency system. Let $\theta \in [0, 1]$ denote the token reward offered by the collector per plastic unit (i.e., bottle), and $P(\theta)$ denote the pickups of plastic items by individuals, which is random following any arbitrary pdf $g_\theta(y)$ and satisfying $g_\theta(0)=0$ by convention. The total token-induced pickup volume is expressed as

$$\boldsymbol{P} = p_0 + \boldsymbol{P}(\theta) \tag{1}$$

The community owns its dedicated cleaning team with a constant capacity. After the token-motivated picking up, the dedicated team needs to clean up the remaining plastic units. Let $\boldsymbol{K} > 0$ denotes the pickup capacity by the dedicated cleaners. Consequently, the ESG utility function is composed of three terms (1) environmental damage, (2) social welfare, and (3) governance cost (Ryberg et al. 2019). The total ESG unitality $\Pi(\theta,e)$ is expressed as

$$\Pi(\theta,e) = \underbrace{\boldsymbol{w}\cdot\min(\boldsymbol{D},\boldsymbol{P})}_{\text{Social welfare}} \underbrace{-\theta\cdot\min(\boldsymbol{D},\boldsymbol{P}) - \boldsymbol{r}\cdot\min\left[(\boldsymbol{D}-\boldsymbol{P})^+,\ \boldsymbol{K}\right]}_{\text{Governance cost}} \underbrace{-\boldsymbol{\xi}\cdot\left[(\boldsymbol{D}-\boldsymbol{P})^+ - \boldsymbol{K}\right]^+}_{\text{Environmental damage}} \tag{2}$$

With some basic algebra, the above can be further written as

$\Pi(\theta,e) = (\boldsymbol{w}-\theta)\cdot\boldsymbol{D} - (\boldsymbol{w}-\theta+\boldsymbol{r})\cdot(\boldsymbol{D}-\boldsymbol{P})^+ + (\boldsymbol{r}-\boldsymbol{\xi})\cdot(\boldsymbol{D}-\boldsymbol{P}-\boldsymbol{K})^+$.

Consequently, the total expected ESG unitality is expressed as

$$\pi(\theta,e) = (\boldsymbol{w}-\theta)\cdot\mathbb{E}\boldsymbol{D} - (\boldsymbol{w}-\theta+\boldsymbol{r})\cdot\mathbb{E}(\boldsymbol{D}-\boldsymbol{P})^+ + (\boldsymbol{r}-\boldsymbol{\xi})\cdot\mathbb{E}(\boldsymbol{D}-\boldsymbol{P}-\boldsymbol{K})^+ \tag{3}$$

For any given enforcement level $\boldsymbol{e}\in[0,1]$, the collector needs to decide the optimal $\theta^*\in[0,1]$ such that

$$\theta^*(e) = \arg\max_{\theta\in[0,1]}\left\{(\boldsymbol{w}-\theta)\cdot\mathbb{E}\boldsymbol{D} - (\boldsymbol{w}-\theta+\boldsymbol{r})\cdot\mathbb{E}(\boldsymbol{D}-\boldsymbol{P})^+ + (\boldsymbol{r}-\boldsymbol{\xi})\cdot\mathbb{E}(\boldsymbol{D}-\boldsymbol{P}-\boldsymbol{K})^+\right\} \tag{4}$$

where the expectations are taken with respect to both random volumes of $\boldsymbol{D}$ and $\boldsymbol{P}$.

The governance supervises the professional collection team with an effort $e\in[0,\ 1]$ subject to the enforcement cost $\phi(e)\geq 0$. Its decision of $\boldsymbol{e}^*\in[0,1]$ is made based on total realized benefit $\pi(\theta^*,e)-\phi(e)$, which is formulated as:

$$e^* = \arg\max_{e\in[0,\ 1]}\left\{\pi(\theta^*,\ e)\ -\ \phi(e)\right\} \tag{5}$$

Assumption 1. The disposing volume of plastic waste follows a normal distribution such that $\boldsymbol{D}(\boldsymbol{e})\sim\mathbb{N}(\mu_e,\ \sigma_e^2)$ and the social-responsible plastic pickers are also following a normal distribution, such that $\boldsymbol{P}(\theta)\sim\mathbb{N}(\mu_\theta,\ \sigma_\theta^2)$.

Lemma 1. Under Assumption 1, we have $\boldsymbol{D}-\boldsymbol{P} = \boldsymbol{D}(\boldsymbol{e}) - \boldsymbol{p}_0 - \boldsymbol{P}(\theta) \sim \mathbb{N}(\mu_e-\mu_\theta-\boldsymbol{p}_0,\ \sigma^2)$, where $\sigma^2 = \sigma_e^2+\sigma_\theta^2$.

Futurework: Under Assumption 1, we shall dive further to glean more insightful results and takeaways that can be used to guide us in making the optimal strategic, operational, and tactic decisions.

Upstream Plastic Manufacturing: Production to Fulfill the Socially Responsible (SR)-Demand

A plastic manufacturing firm designs, assembles, and sells a plastic-contented product (e.g., plastic bottles or packages) over a horizon $\mathcal{T}$ consisting of $T>0$ periods (e.g., in an annual basis) where the time period t as 1, 2, …, T. The product is made of an amount $\alpha \in [0,1]$ of key recycled plastic and $1-\alpha$ virgin plastic. We refer to $\alpha \in [0,1]$] as the *recycling grade*. Regulatorily, there is a typical mandatary standard to enforce the recycling grade above a specific level. On the demand side, SR demand is a crucial market player. Hence, the product grade induces a stochastic SR demand $\boldsymbol{D_t}(\alpha)$ in each period $\boldsymbol{t}$. On the demand side, to reflect the fact that the recycled plastic material greatly contributes to attracting more demand for its product, we simply assume that the higher the grade, α, the larger the demand (in the stochastic sense). We also assume that virgin plastic material is always available at a unit cost of $c_{\boldsymbol{v}}$ on the supply side. The virgin plastic material is always available from a spot market, and the delivery lead time is ignorable. For instance, global fossil fuel-based plastics production is dominated by large petrochemical companies, including some major oil and gas producers. In contrast, recycled material is subject to limited availability $\boldsymbol{S}_t$ during the period $\boldsymbol{t}$. Importantly, it poses a high price of $c_{\boldsymbol{r}}$ per unit such that $c_{\boldsymbol{r}} > c_{\boldsymbol{v}}$, thus sourcing recycled plastic material must be managed carefully.To address the aforementioned plastic production challenge, we develop a Dynamic Programming (D.P.) model. At the beginning of each period $\boldsymbol{t}$, the manufacturer updates the total product units as x_t. While anticipating random demand $\boldsymbol{Z}_t$, the manufacturer needs to decide how much virgin plastic s_v and how much recycled plastic s_r for this period's production. In this case, the total profit can be formulated as

$$\Phi_t(x_t, s_r, s_v, \boldsymbol{Z}_t) = \underbrace{\boldsymbol{p}_t \cdot \min\left(\boldsymbol{Z}_t, \boldsymbol{I}_t\right)}_{\text{Revenue}} - \underbrace{\left(\boldsymbol{c}_r \boldsymbol{s}_r + \boldsymbol{c}_v \boldsymbol{s}_v\right)}_{\text{Sourcing cost}} \underbrace{- \boldsymbol{h}_t \left(\boldsymbol{I}_t - \boldsymbol{Z}_t\right)^+ - \boldsymbol{b}_t \left(\boldsymbol{I}_t - \boldsymbol{Z}_t\right)^-}_{\text{Inventory Cost}} + \underbrace{\gamma_t \cdot \psi_{t+1}(\boldsymbol{I}_t - \boldsymbol{Z}_t)}_{\text{Profit-to-go}} \quad (6)$$

where $\boldsymbol{I}_t = x_t + s_r + s_v$ denotes the total volume of the product after production, and it is the sum of the beginning inventory level x_t and the new production quantity. Financially, $\gamma_t \leq 1$ is the discount factor reflects the monetary value of time (this can be obtained by the interest rate). For the period $\boldsymbol{t}$, given the supply available of recycled plastic of $\boldsymbol{S}_t$, the optimal sourcing and production are jointly determined dynamically by

$$\psi_t(x_t) = \max_{\alpha \cdot (s_r + s_v) \leq s_r \leq S_t} \mathbb{E}\left[\Phi_t(x_t, s_r, s_v, \boldsymbol{Z}_t)\right], \quad (7)$$

where the expectation is taken with respect to the random demand $\boldsymbol{Z}_t$, and the constraint $\alpha \cdot \left(s_r + s_v\right) \leq s_r$ ensures the regulation of at least α recycling grade. In this dynamic programming model, the terminal condition is $\psi_{T+1}(y) = 0$.

Smart Contracts can be adopted and implemented as a solution key to address the supply shortage and uncertainty. As raw plastic (mainly recycled plastic) is of uncertain supply and volatile market cost, sourcing and procurement become critical to the manufacturers. In order to mitigate the supply risk (i.e., supply shortage and varying price), the firm typically leverages a multi-sourcing strategy and sources the green material from a pool of M suppliers. Each supplier m is characterized by a different supply capacity and supply cost per unit. In this case, making the optimal procurement decision to support production is strategically essential. As introduced above, Smart Contract is a practice of Blockchain technology, and it provides automatic execution of a real-time contract. Therefore, the plastic manufacturer can always secure the best supply of recycled plastic materials in this fashion.

Closing the Loop: Connecting the Up- and Down-streams

In the supply chain under consideration, the finished plastic product (e.g., plastic bags, packaging containers), there will be $\boldsymbol{\lambda} \in [0,100\%]$ a portion of the product goes to the ecosystem under study (cf. Figure 2). That close-looped plastic can be tracked and traced through the Blockchain platform, which will build up the sustainability and equilibrium of the CE. Among those products, the disposing volume $\boldsymbol{D}$ will be a random variable determined by $\boldsymbol{\lambda}$.

Future work: Solve the dynamic programming model and develop some efficient and effective computational algorithms. Importantly, through the solution approach, we shall glean useful and actionable recommendations to the recycling supply chain. For closing the loop, we shall study the optimal $\boldsymbol{\lambda}^*$ such that the whole ecosystem stands with an equilibrium balance. Based on the obtained equilibrium solution, we shall gauge the effectiveness of the Blockchain platform. In particular, we shall visualize how much economic impact we can gain from the Blockchain technology adoption.

Social-Responsible Behavior

Uncertainty from the market and consumer behavior is a critical challenge in the CE system. One of the major uncertainties encountered in the recycling system is the stochastic demand process for recycled-plastic products. The other uncertainty is the pick-up behavior of individuals incentivized by Blockchain-platform token reward or cryptocurrency.

Sustainability of the Plastic CE

The model and analysis can be readily applied for the production/inventory systems of a perishable product with a constant shelf-life $\boldsymbol{T}>0$. The inventory is depleted according to the descending age of the products that have not perished. Hence, the maximum availability of the valid product is $\boldsymbol{\ell} = \boldsymbol{\rho} \cdot \boldsymbol{T}$, observing that the product staying in the system longer than $\boldsymbol{T}$ (if any) becomes perished and is removed from the inventory system. Therefore, the inventory process is the same as that for perishable products with a lifetime $\boldsymbol{T}$. In this case, this research also embraces the study of perishable products.

Production Capacity Management

Capacity management exposes another concern in the production/inventory system. How to determine the proper capacity level of the inventory $\boldsymbol{\ell}$ becomes imperative. In this case, the optimal solution $\boldsymbol{\ell}$ sheds light on strategically selecting the inventory capacity level. The analytical result on the optimal $\boldsymbol{\ell}$ sheds light on capacity management.

DISCUSSION AND CONCLUSION

The increasing use of plastic has led to a global plastic crisis, with the vast amount of plastics polluting the environment and threatening the ecosystem. The global concern over plastic waste and pollution has resonated with the public, corporations, and governments alike. Unfortunately, dealing with the aftermath of the plastic crisis via traditional waste management systems is insufficient and inadequate in the plastic crisis. A survey conducted in the year 2022, which included 1028 respondents from the U.S., reported that plastic recycling in the U.S. needs improvement. The respondents suggested that there needs to be a shift in Americans' attitude toward plastic disposal (Statista, 2022). At the same time, imposing a tax on WMF for waste disposal could deter the firm from engaging in waste processing activities (Chintapalli & Vakharia, 2023).

This study leverages data and technology to propose a model to tackle the pressing issue of plastic pollution. In particular, it employs BCT to enhance the efficiency of plastic recycling. This study may serve as a guideline for regulatory bodies and policymakers. Reusing, recycling, and remanufacturing plastic within the value chain promotes a more sustainable, eco-friendly ecosystem. The literature discusses four central challenges of incorporating BCT: Policies and Regulations, Materials and Resources, Technology and Knowledge, and the training of Users. Finally, a paradigm shift is needed from the current linear consumption model to a more circular and sustainable approach that fosters reducing, reusing, and recycling of plastics. Therefore, it is essential to employ collective efforts to streamline the recycling process by enhancing the recycling facilities, advocating for public participation in plastics collection, minimizing contamination, and boosting the market interest in recycled over virgin plastic.

REFERENCES

Ahmad, R. W., Salah, K., Jayaraman, R., Yaqoob, I., & Omar, M. (2021). Blockchain for waste management in smart cities: A survey. IEEE Access, 9, 131520-131541.

Bing, X., Bloemhof-Ruwaard, J., & Van der Vorst, J. (2014). Sustainable reverse logistics network design for household plastic waste. Flexible Services and Manufacturing Journal, 26.

Bhubalan, K., Tamothran, A. M., Kee, S. H., Foong, S. Y., Lam, S. S., Ganeson, K., ... & Ramakrishna, S. (2022). Leveraging blockchain concepts as watermarkers of plastics for sustainable waste management in progressing circular economy. Environmental Research, 213, 113631.

Chang, A., El-Rayes, N., & Shi, J. (2022). Blockchain technology for supply chain management: A comprehensive review. FinTech, 1(2), 191-205.

Chang, A.; Katehakis, M.N.; Melamed, B.; Shi, J. (2021). Blockchain Technology and its Applications, Working paper, NJIT.

Chaudhuri, A., Subramanian, N., & Dora, M. (2022). Circular economy and digital capabilities of SMEs for providing value to customers: Combined resource-based view and ambidexterity perspective. Journal of Business Research, 142, 32-44.

Chaudhari, U. S., Johnson, A. T., Reck, B. K., Handler, R. M., Thompson, V. S., Hartley, D. S., ... & Shonnard, D. (2022). Material Flow Analysis and Life Cycle Assessment of Polyethylene Terephthalate and Polyolefin Plastics Supply Chains in the United States. ACS Sustainable Chemistry & Engineering, 10(39), 13145-13155.

Chen Z, Yildizbasi A, Sarkis J (2023) How safe is the circular economy?. Resources, Conservation and Recycling, 188, 106649.

Chidepatil, A., Bindra, P., Kulkarni, D., Qazi, M., Kshirsagar, M., & Sankaran, K. (2020). From trash to cash: how blockchain and multi-sensor-driven artificial intelligence can transform the circular economy of plastic waste? Administrative Sciences, 10(2), 23.

Chintapalli, P., & Vakharia, A. J. (2023). The waste management supply chain: A decision framework. Decision Sciences.

Daily, J. (2021). Mathematically Modeling Plastic Pollution in the Great Lakes [Ph.D. Thesis]. Rochester Institute of Technology.

El-Rayes, N., Chang, A., & Shi, J. (2023). Plastic Management and Sustainability: A Data-Driven Study. Sustainability, 15(9), 7181.

El-Rayes, N., Chang, A., & Shi, J. (2022). Plastic Crisis: A Business Data-Driven Study on the Evolution and Comparison of its Practices and Research. Proceedings of the 51st Northeast Decision Sciences Institute (NEDSI) Annual Meeting. New Jersey, 371-396.

El-Rayes, N., & Shi, J. (2022). A Paradigm Shift for Addressing the Plastic Crisis: A Data-Driven Study. Proceedings of the 53rd Decision Sciences Institute (DSI) Annual Conference. Texas, 70-82.

Environmental Protection Agency (EPA). (2018). Containers and Packaging: Product-Specific Data. https://www.epa.gov/

Forbes. (2020). Guess Which Two Countries Produce The Most Plastic Trash Per Person?. https://www.forbes.com

Fossi, M. C., Romeo, T., Baini, M., Panti, C., Marsili, L., Campani, T., Canese, S., Galgani, F., Druon, J.-N., & Airoldi, S. (2017). Plastic debris occurrence, convergence areas and fin whales feeding ground in the Mediterranean marine protected area Pelagos sanctuary: A modeling approach. Frontiers in Marine Science, 167.

Geyer, R., Jambeck, J. R., & Law, K. L. (2017). Production, use, and the fate of all plastics ever made. Science advances, 3(7), e1700782.

Gong, Y., Wang, Y., Frei, R., Wang, B., & Zhao, C. (2022). Blockchain application in circular marine plastic debris management. Industrial Marketing Management, 102, 164-176.

Guyot Phung, C. (2019). Implications of the circular economy and digital transition on skills and green jobs in the plastics industry. Field Actions Science Reports. The Journal of Field Actions, (Special Issue 19), 100-107.

Hardesty, B. D., Harari, J., Isobe, A., Lebreton, L., Maximenko, N., Potemra, J., Van Sebille, E., Vethaak, A. D., & Wilcox, C. (2017). Using numerical model simulations to improve the understanding of micro-plastic distribution and pathways in the marine environment. Frontiers in Marine Science, 4, 30.

Hristova, T. (2022, June). The place of the blockchain in the recycling of raw materials. In 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE) (pp. 1-4). IEEE.

Katz, D. (2019). Plastic bank: launching social plastic® revolution. Field Actions Science Reports. The Journal of Field Actions, (Special Issue 19), 96-99.

Khadke, S., Gupta, P., Rachakunta, S., Mahata, C., Dawn, S., Sharma, M., Verma, D., Pradhan, A., Krishna, A. M. S., & Ramakrishna, S. (2021). Efficient plastic recycling and remolding circular economy using the technology of trust–blockchain. Sustainability, 13(16), 9142.

Kouhizadeh, M., Zhu, Q., Sarkis J (2020) Blockchain and the circular economy: Potential tensions and critical reflections from practice. Production Planning and Control, 31(11- 12), 950–966.
Mougayar, W. A decision tree for blockchain applications: Problems, opportunities or capabilities. Startup Manag. 2015, 3, 23–43.

Kumar, A., Liu, R., & Shan, Z. (2020). Is blockchain a silver bullet for supply chain management? Technical challenges and research opportunities. Decision Sciences, 51(1), 8-37.

Lau, W. W., Shiran, Y., Bailey, R. M., Cook, E., Stuchtey, M. R., Koskella, J., Velis, C. A., Godfrey, L., Boucher, J., & Murphy, M. B. (2020). Evaluating scenarios toward zero plastic pollution. Science, 369(6510), 1455–1461.

Nöbauer, M. (2022). Foster Trust in Industrial Quality Control by Introducing Smart Contracts. Proceedings of the 15th International Conference Information Systems (IADIS).

El-Rayes, N., & Shi, J. (2022). A Paradigm Shift for Addressing the Plastic Crisis: A Data-Driven Study. Proceedings of the 53rd Decision Sciences Institute (DSI) Annual Conference. Texas, 70-82.

Rejeb, A.; Rejeb, K.; Simske, S.; Treiblmaier, H. Blockchain Technologies in Logistics and Supply Chain Management: A Bibliometric Review. Logistics 2021, 5, 72.
Saberi S, Kouhizadeh M, Sarkis J, Shen L (2019) Blockchain technology and its relationships to sustainable supply chain management. International Journal of Production Research, 57(7), 2117-2135.

Ryberg, M. W., Hauschild, M. Z., Wang, F., Averous-Monnery, S., & Laurent, A. (2019). Global environmental losses of plastics across their value chains. Resources, Conservation and Recycling, 151, 104459.

Sarkis J, Helms M, Hervani A (2010). Reverse logistics and social sustainability. Corporate social responsibility and environmental management, 17(6), 337-35

Statista.(2022). Plastic Waste in the U.S. [Industries & Markets]. https://www.statista.com/

Walden, S. (2020). Revamping the plastics cycle with emerging technologies. Retrieved from https://www.dell.com/en-us/perspectives/plastics-rethinking-the-cycle/, August 1 2022.

DECISION SCIENCES INSTITUTE

Give Me a Choice! A Field Experiment on Autonomy in a Cyber-Physical Production System

Daniel Kwasnitschka
Chair of Production and Operations Management, D-MTEC, ETH Zurich, Switzerland
Email: dkwasnitschka@ethz.ch

Henrik Franke
Faculty of Management, Economics and Social Sciences, University of Cologne, Germany
Email: franke@wiso.uni-koeln.de

Torbjørn Netland
Chair of Production and Operations Management, D-MTEC, ETH Zurich, Switzerland
Email: tnetland@ethz.ch

ABSTRACT

This paper reports on a field experiment in a pioneering cyber-physical production system using wearable devices. We investigate the impact of allowing workers to autonomously choose their tasks through smartwatches. The study draws on data from a 12-week experiment in two European manufacturing plants, analyzing 65,348 machine status reports and 49,050 work tasks. Results show that while this approach improves production efficiency by 6.9%, it unexpectedly slows workers' response to new tasks by 73 seconds, with no significant effect on task completion speed. These findings challenge existing theories on human motivation and provide implications for managers designing advanced production systems.

KEYWORDS: Field experiment, advanced manufacturing technology, autonomy

INTRODUCTION

Manufacturing is one of the most important industrial sectors on the globe. It contributes 27% to the global GDP and shows a growth trend since the financial crisis in 2009 (World Bank, 2023). Thus, adequately managing manufacturing operations is of paramount importance and so is understanding the advancing digital technologies that promise to make long-standing challenges of manufacturing management obsolete. Several examples in the literature focus on relatively clearly defined problems with relatively narrow scope —a use case (Maghazei et al., 2022)—such as dealing with variability in quality management using artificial intelligence (Senoner et al., 2022) or using 3D printing in spare parts manufacturing (Baumers & Holweg, 2019; Heinen & Hoberg, 2019). However, we lack studies that trace effects of novel technologies on entire production systems. This is an important omission since interconnectedness of various technologies is one of the essential elements of the fourth industrial revolution (OECD, 2017). Its revolutionary character does not derive from novel technology alone, like for the three previous revolutions, but from creating internet-of-things, which are networked structures of mutually communicating technology. This study is one of the first to examine how changes in a pioneering cyber-physical production system enabled by novel technologies affect workers' behavior and performance. It derives conclusions on how to configure interconnected production systems in harmony with human workers.

Managers in production processes that rely on automation commonly assign workers to areas of the shop floor to monitor the machines located within them. This is a commonplace heuristic to solve an ever-present challenge of production systems: allocating the available workforce to pending production tasks with the aim of improving overall efficiency. This problem is challenging to solve when managers lack accurate information regarding the current availability of workers, the status of tasks, and requirements that tasks may have in terms of skills. Digitalization has reduced these information gaps considerably. For example, new internet-of-things applications can generate alerts about unplanned events and provide the real-time availability of personnel and their skill profiles. These developments enable the use of new technologies for matching tasks and workers to achieve a fundamentally new way of production organization.

The production system we analyze has changed the common organization of shop floors by areas to an organization by skills by using an internet-of-things and industrial smartwatches. By codifying workers' skills and all possible tasks, matching skills and tasks, and connecting machines and workers through smartwatches, the system makes possible operating production without a division into areas. These new ways of organizing uncover new problems with scientific relevance (Holmström et al., 2019). Should workers be assigned to a task or choose among all pending tasks? We study the effect of giving workers the ability to autonomously choose their tasks in a formerly controlling cyber-physical production system in this study. The a-priori implemented system delegates production tasks to available and adequately skilled shop floor workers once a task, such as a machine interruption, has been detected by sensors in real time. Both variants only show those tasks to a worker that match their individual skills.

The a-priori implemented production system based on delegation should radically reduce the need for management and coordination by directing available workers to tasks that match their skills seamlessly. This should increase the overall performance of the plant as we know that sharing resources across tasks is efficient (Jordan & Graves, 1995). However, human agency matters, and humans are not machines (Boudreau et al., 2003; Donohue et al., 2020; Shunko et al., 2018). The way operators perceive, use, and help improve the system ultimately determines whether the system will increase performance. Thus, we are interested in the behavioral side of the novel production system. Specifically, we hypothesize that the way that workers receive their next job task plays an important role in the system's overall performance. We leverage theory on fundamental needs in psychology (Deci & Ryan, 1980) to hypothesize that performance is higher when workers can act autonomously and choose their next task instead of receiving one from the system.

We compare the performance of two alternative systems in a field experiment among two digitalized factories with real operators working in real manufacturing operations. We analyze 65,588 machine status reports and 59,050 work tasks completed in two manufacturing plants of a Tier supplier of the automotive industry in Germany and Italy to examine whether offering a choice improves performance. We compare the performance before and after the treatment and compare the results with a control group in a different plant of the same company using a Difference-in-Difference (DiD) approach. In addition to overall production efficiency, we also observe behavioral effects on workers' task responsiveness (time until pending tasks are accepted) and task efficiency (time required for task completion). We find that the manipulated system increases production efficiency by 6.9% (95% confidence interval ranges from 4.5% to 9.3%). However, against our assumptions, it also slows workers' responsiveness by 73 seconds (95% confidence interval ranges from 47.6 sec to 98.8 sec) and does not significantly affect how fast workers complete tasks. The discussion around these interesting effects guides scholars and managers on how to configure human–machine interfaces of cyber-physical production systems. We also contribute to the discussion of autonomy, an important variable of fundamental theories of human motivation, and its ambivalent nature in a pioneering digitalized manufacturing context.

THEORETICAL BACKGROUND

Literature Review

Distributing tasks among workers is central to management sciences and operations management. Past research has frequently studied task selection and assignment from an operations research perspective by modeling task assignment (Ernst et al., 2006; Lewis et al., 1998), task reassignment and worker collaboration (Işik et al., 2021), flexible versus fixed assignment systems (Zavadlav et al., 1996), or dynamic task assignment protocols (Lim & Wu, 2014). While these valuable contributions have made important headway into research on task assignment, they assume that workers do not interpret their work assignments in terms of their own agency but proceed directly to work completion. However, humans are not machines (Shunko et al., 2018). We extend prior research by considering worker behavior in our research on task assignment in production. Furthermore, prior studies have not addressed shifts from one system to another or measured the resulting effects on worker behavior. Thus, the behavioral and socio-technical effects of such changes in production systems—for instance, from low autonomy task delegation to high autonomy task selection—are underrepresented in the current body of literature. At the same time, studies have indicated that worker autonomy is one of the most important work organization variables affected by smart technologies in manufacturing (Cagliano et al., 2019).

Despite its relevance, only relatively few scholars have so far focused on workers' autonomy in task selection whereas, for example, the literature is more outspoken about autonomy in teams (Langfred, 2005; Langfred & Moye, 2004), autonomous choices of consumers (Basu & Savani, 2017), or the autonomy of organizational units (Arregle et al., 2023). A relatively rare example focusing on worker autonomy is Raveendran et al. (2021), who suggested that self-selection outperforms traditional delegation of tasks to workers by managers. They found that self-selection has an advantage, particularly when workers have special skills but narrow responsibilities, tasks are modular, and difficult to plan. In addition, Baldwin and Clark (2006) have showed that when task division is transparent, self-selected work increases the likelihood that workers will choose to contribute rather than free-ride. Moreover, the laboratory experiment in Kamei and Markussen (2022) has shown that democratic task assignment can improve efficiency in work teams. Furthermore, the autonomy of choosing also entails reviewing different options. In that vein, Basu and Savani (2017) studied the effect that option presentation has on decision quality. The study found that individuals make better decisions and engage in deeper cognitive processing when they view options simultaneously rather than viewing them sequentially.

All of this relevant literature has informed management but focused on contexts that are different from behavioral production management. Raveendran et al. (2021) focused on an agent-based model that ignores possible behavioral influences rooted in autonomy, Baldwin and Clark (2006) researches software development, and Basu and Savani (2017) examine a choice between options of different consumer products. Thus, manufacturing remains under-represented despite its immense relevance. It remains unclear if these results hold for other types of decisions—for instance, decisions on open work tasks in a production setting that are taken at a much higher rate. In that setting, we propose that workers change their behavior when they are granted the autonomy to review and choose tasks. Specifically, our study attempts to show that autonomy enabling freedom of choice increases workers' efforts and can increase production efficiency. Thereby, we directly continue the current discussion on choice at work from Kamei and Markussen (2022). Our study extends the published research with field evidence on performance differences by enabling workers to autonomously choose tasks over several weeks in real factories that produce millions of parts per year.

Autonomy and Human Motivation

Autonomy is the capacity to self-organize one's experience and behavior (Deci & Ryan, 2000). According to the self-determination theory on human motivation, autonomy is a basic requirement for the well-being and growth of an individual (Deci & Ryan, 2000). More specifically, when an individual feels autonomous and in control of their outcomes, their innate need for autonomy is satisfied, and intrinsic motivation is enhanced. On the other hand, when individuals experience the surrounding environment as controlling, intrinsic motivation decreases (Deci et al., 1989). The intrinsic perspective fits our study context as workers receive no additional pay per completed task at our field partner. Furthermore, the theory posits that supporting autonomy will yield other adaptive outcomes, such as improved performance, well-being, and psychological growth (Deci & Ryan, 2000; Patall et al., 2008). Studies have further distinguished several facets of autonomy, such as the autonomy to choose one's methods at work, to schedule one's work, and to choose evaluation criteria (Breaugh, 1985). Our study relates most to work scheduling autonomy, which is the extent to which workers can sequence their work activities (Breaugh, 1985).

Self-determination theory has a rich history in management and organizational research (Deci et al., 2017; Van den Broeck et al., 2016) but somewhat less in operations and production management. However, the behavioral autonomy concept and self-determination theory are not entirely new to operations management. The concepts has been leveraged to explain how to best request workers for inventory auditing tasks via a crowd-sourcing platform (Ta et al., 2021) or differences in the performance of call center employees that face fluctuating workload (Ashkanani et al., 2022). Other studies relate less to operations or process management and emphasize the predictive power of the theory in supplier selection (Roehrich et al., 2017) or in project management (Sarafan et al., 2022), for instance. In our context, autonomous choice is relevant, since whether and how workers can choose tasks enabled by an internet-of-things production system has a clear connection to the idea of self-organization that defines autonomy (Deci & Ryan, 2000). Consistently, related studies also point to self-determination theory (Kamei & Markussen, 2022) and it has been suggested that production digitalization can lead to employee empowerment in production (Leyer et al., 2019), for instance by increasing their control over tasks' structure, frequency, or priorities (e.g., Kembro et al., 2017; Sharma et al., 2019). However, digital systems in practice still often turn out to be a controlling feature of shop floor employees' work day. We study the effect that increasing worker autonomy via choice has in a manufacturing context.

Generally, the provision of choice is one of the most obvious ways to enhance an individual's experience of autonomy (Patall et al., 2008). Regarding this, self-determination theory generally predicts that autonomy should result in positive performance outcomes (Deci & Ryan, 1980, 1985, 2000) but few studies show that choice, as a form of autonomy, drives performance. Thus, despite the scholarly tradition of the theory and choice as a generic concept, the aim of our study is novel since, first, the literature has taken other directions, and, second, we test the effect of autonomous choice enabled by technology that was hitherto unavailable. Specifically, the few studies on the issue focus on consumer choices or research human behavior via computation (Basu & Savani, 2017; Raveendran et al., 2021). We extend this stream into the direction of manufacturing and offer real-world data. Other closely related research has factored in the disappointment that results when a system ignores a worker' earlier solicited preference (Kamei & Markussen, 2022). We analyze choice alone in this study. Furthermore, our study is novel as digitalization changes the inherent assumptions of production management. Specifically, technology made all production problems visible in near real time, which was unheard of before digitalization. An advanced system allows organizing fast-paced and highly granular production tasks by activities and skills instead of areas, as discussed above, which has hitherto only been documented for white-collar work (Bals & Turkulainen, 2017), to the best of our knowledge. Thus, we address a gap that has been opened by radical change induced by technological development. Our research setting is a rare example of such pioneering technology.

Theoretical Predictions

Our study proposes differences between two distinct production systems that we refer to as *delegation* and *choice*. *Delegation* means that a digital system automatically matches open tasks and workers by assigning workers one specific task that they are expected to complete. This means that workers have no *choice* among tasks and thus little autonomy in the sense of self-determination theory. At the same time, workers cannot easily see what other open tasks are available. In the case of *choice*, workers choose a task from a list of available tasks. Due to their ability to choose, their autonomy is high in the sense of self-determination theory.
We argue that the transition from *delegation* to *choice* will increase production efficiency. Production efficiency is the ratio between actually produced units and the expected production volume. The expected volume is the design capacity considering planned downtimes such as setups. We argue that increased autonomy will create benefits for workers' intrinsic motivation, as they feel more in control and self-efficacious in a production system that informs them about all pending tasks. Workers perceive a production systems' user interface that provides rich information as a supporting element provided by their organization.

Similar motivational effects have been documented for autonomy of the workforce (Langfred & Moye, 2004). Furthermore, it has been indicated that higher autonomy also increases the breadth of activities that employees complete (Morgeson et al., 2005). This means that workers will not only respond with higher effort in the *choice* scenario but also address a wider range of problems as before, which creates pooling benefits as interruptions can be addressed by a workforce that uses overlapping roles and skills. Generally, workers will likely increase their efforts as a way to reciprocate the provision of *choice* in the *choice* scenario. Consistently, research has shown that reciprocation is an important mechanism to increase performance by offering more support to workers, such as via autonomy (Eisenberger et al., 2001). We expect that reciprocation will increase the production efficiency of the shop floor. Thus, we hypothesize:

Hypothesis 1. Increasing workers' ability to autonomously choose tasks ("choice") will increase production efficiency relative to delegating task directly to them ("delegation").

Self-determination theory has been suggested to also hold implications for task performance of individuals (Williams & Luthans, 1992). In our context of digitalized production, we conceptualize task performance as task responsiveness and task efficiency. Task responsiveness denotes the time that a task is waiting for processing until an idle worker accepts it. Task efficiency denotes the time required for completing the task once accepted. Research provides some indications that workers who choose tasks of their personal preference exert more effort in general (Iyengar & Lepper, 2000; Patall et al., 2008), employees show more prosocial behaviors when granted autonomy (Donald et al., 2021), and autonomy increases the effect of useful personality traits on job performance (Barrick & Mount, 1993). We add a perspective of skill to this with our arguments.

In our context, workers generally only see those tasks on the smartwatch for which they have been assigned the codified skill. Yet, skill is the system's binary codification of true skill, which no production system truly knows. Workers have tacit knowledge on what codified skills and tasks they have the highest mastery of. This discrepancy is natural and allows workers to pick tasks for which they have a personal preference among the ones they are able to complete as per the codified skill. These preferred tasks are enjoyable or interesting to workers, and one can therefore expect them to apply themselves to the task with higher intensity, which decreases the processing time. Furthermore, workers enjoy applying their best developed skills as it can trigger a feeling of flow and flourishing and will pick these tasks first (Kamei & Markussen, 2022; Spreitzer et al., 2005). This matches true skills and task requirements, such that task efficiency increases. We expect that mechanisms will reduce the required time for solving a machine interruption on the shop floor. Accordingly, we hypothesize:

Hypothesis 2. Increasing workers' ability to autonomously choose tasks ("choice") will increase task efficiency relative to delegating task directly to them ("delegation").

We argue that the *choice* scenario will have a positive effect on task responsiveness—that is, reduce the time a task is before acceptance by a worker. Self-determination theory proclaims that individuals who have the freedom to choose develop higher intrinsic motivation for what they do (Patall et al., 2008). When individuals feel autonomous and are more intrinsically motivated, they proactively work on tasks they see (Grant et al., 2011). Once workers have the autonomy to choose among a transparent set of options, the positive effects can show in terms of pro-active behavior, which includes faster responses to pending work tasks. This is consistent with studies that have shown that autonomy makes pro-active behavior more likely to contribute to task performance features (Kim et al., 2009).

Contrary to the *choice* scenario, pro-activism is not possible in the *delegation* scenario, as the production system dictates specific tasks to workers and expects them to accept. This system actively reduces pro-activism. Consistently, a lack of autonomy can hinder pro-activism in the form of making quick decisions by tempting employees to focus on the controlling situation, which impedes them from taking actions (Bledow et al., 2022). The *choice* scenario has the potential to increase task responsiveness by dismantling controlling features of a production system that can impede workers' pro-activism. As a result, it will be more likely that workers will start working on pending tasks—their preferred tasks—more quickly in the *choice* scenario compared to the *delegation* scenario. Thus, we hypothesize:

Hypothesis 3. Increasing workers' ability to autonomously choose tasks ("choice") will increase task responsiveness relative to delegating task directly to them ("delegation").

FIELD SETTING

Study Design

Our study uses a pre-post assessment of human behavior and system performance before and after a treatment in a controlled field environment (Rea et al., 2021). We use two comparable production facilities of the same company as treatment and control group. Our experiment is a quasi-experiment as the allocation of the treatment had to be agreed upon by management. We verify the similarity of the groups in detail and pick this design for its several strengths: Using a field experiment allows us to test our hypotheses in a real production environment that is ideal in terms of the experiments' ecological validity and realism (Morales et al., 2017). Second, our participants are real production workers ("targeted sample") instead of alternative participant sources, such as student pools, which can constitute a limitation for research designs (Eckerd et al., 2021). Third, all actions and decisions taken in our design by participants are consequential (i.e., not hypothetical), and we observe real operational outcome variables, which are both recommended features of rigorous experiments (Bachrach & Bendoly, 2011).

We ran our study in collaboration with a global manufacturing firm that produces complex metal parts as a Tier supplier for the automotive industry. We focus on the automotive industry since it contributes about 3% of the global economic output alone and on a large supplier (ca. 9 bn USD annual turnover) since suppliers are a major source of technological advancement in that industry (Ewing & Cohen, 2021; Jean et al., 2014). The participating plants are located in central Europe and produce around 500 thousand parts per day in a three-step process, including physical and thermal treatment. All production steps are heavily automated using robotics to move parts into and out of the presses. Parts receive thermal treatments in few but large oven-type facilities where interruptions are easily detected. Therefore, the faster-paced and batch-production step of pressing/molding is a more appropriate environment for our study. The molding production area encompasses approximately 50 machines per plant that operate in three main status: producing parts, planned downtime like setup, or unplanned downtime, which are mostly interruptions. Interruptions can occur due to a lack of raw material, full finished goods buffer, or technical malfunctions, such as a collision of the robotics or a damaged tool. We focus on these machine interruptions as they represent a significant source of machine downtime in the factory.

Workers do not roam production to spontaneously find and solve problems as they would in non-digitalized production environments. Instead, machines notify a central system via sensors that matches tasks to available and adequately skilled workers in real time. The system does not prioritize among the workers and notifies all available workers simultaneously by directly prompting them on their wearable smartwatch (*delegation*). When a worker accepts a task, the task disappears from others' smartwatches and they do not receive additional prompts until completion.

Treatment

The treatment involves changing the work allocation for the treatment group from showing unsolicited prompts on workers' smartwatches (*delegation*) to displaying a list of open tasks to actively choose from (*choice*). Each worker's list, just as in the prior system, only shows tasks that match their codified skills and one task can appear on several lists at the same time. Thus, the assignment process remained unchanged in the experiment besides the treatment illustrated in Figure 1.

Figure 1: Overview of delegation (left) and choice (right)

On the one hand, the left picture illustrates the *delegation* scenario. The smartwatch shows a push notification about an interruption at palletizer 3175 at 15:46. The interruption is described as a blockage of the chain conveyor. In this scenario, workers see one task at a time and face pressure to accept. Rejection is technically possible, but acceptance is generally expected. This represents low autonomy. On the other hand, the right picture shows the *choice* scenario. The smartwatch displays a list of pending production tasks that match the skill profile of this worker. The upper is an interruption at machine 2321, the lower an interruption of palletizer 4187. This reflects high autonomy.

Procedure

Our experimental procedure included two main phases. In the first phase, production was carried out with the initial *delegation* system. This system had been implemented months ago. To reduce the impact of exogeneous influences, we reduced the period that we considered in our analyses to the two weeks before the system change. The second phase encompassed the following two weeks after the system change to allow a comparison of the two periods and to further minimize any exogenous effects. We ran sensitivity analyses alternating these choices.

The system change was communicated via established communication channels from management to the shop floor via team leaders, shift meetings, and printed one-page summaries one week before the actual change. On the day of the system change, two researchers attended all the shift handover meetings to assist in answering questions regarding the new system. Thus, our study does not use deception but does not actively disclose the hypotheses of our research (Eckerd et al., 2021). We carefully avoided giving instructions on whether and how the workforce should change its work routines or behavior in response to the change and emphasized that the research team draws no immediate gains from the results. Thus, we minimized demand effects via desirability (Eckerd et al., 2021). Previous studies have demonstrated that only strong deliberate signals could trigger significant demand effects (De Quidt et al., 2018), which we avoided in our procedures. In addition, we reduced the anyhow low power distance between us (i.e., externals to the company hierarchy) and the workers in regular observational visits before the manipulation to reduce Hawthorne effects. In other words, the situation of researchers being present was not new or surprising to workers such that it could trigger the effect. We verify this analytically. We did not pay any incentives during our experiment to avoid altering the established reward structure.

We did not perform a traditional manipulation check as part of our study since our treatment is an objective variation of parameters. Our manipulation fundamentally affected the production processes and left no chance of going unnoticed or being misinterpreted. An additional manipulation check regarding workers' new opportunities in the *choice* system vis-à-vis the *delegation* system seemed unnecessary and a possible source of demand effect. Such biases can emerge when participants are made aware of the central concepts under study and guess the underlying research questions (Eckerd et al., 2021).

EMPIRICAL STRATEGY

Data Strategy

We were able to gather data on all machine interruptions that occurred in the 12-week period (59,050 tasks) and the unit output efficiency at the machine level for the same period (65,588 machine statuses) for the treatment and control group. We analyzed two different types of information for each of the tasks recorded in our data set. Each task was pending for a specific time until acceptance by a worker. This measure represents task responsiveness in our study. Furthermore, each task has a specific time that a worker spent solving the problem. This measure represents task efficiency in our study. In addition, we measured the machine-level production efficiency as the relation between the actual output versus expected output, which varies around 1.0 for under- or over-performance, respectively. The expected output is an internal and machine-specific efficiency benchmark metric that computes an expected output from the design capacity, average setup time, and average troubleshooting times for tasks such as material refills or trolley changes. This benchmark is a fixed value for each machine-type–product combination and is provided by the company. All of these measures are directly observed, as recommended for experiments (Bachrach & Bendoly, 2011).

We included dummy variables to differentiate the three shifts and included the current utilization as a control variable in the regression equation. We calculated the utilization as a workload ratio between the net total number of active machines (i.e., excluding machines in planned downtime) and the clocked hours of the workers.

To ensure high-quality and reliable data, we applied several data processing steps before the analysis with the two data sets we used: the production efficiency data set, taken directly from the manufacturing execution system, and the task performance data set, taken from the smartwatch system back end. In the production efficiency data set, which is the basis for testing H1, we excluded all reports from machines that went into planned or unplanned maintenance. This was important, since we wanted to measure the system performance of different task *delegation* systems for interruptions, not the performance of maintenance. Maintenance was also not assigned via the smartwatch at this time. Moreover, we excluded the weekends and days when the company decided to pause production to make use of federal reduced working hours programs targeted at reducing the economic effects of the pandemic. These are periods that do not represent normal operations. The task performance data set, which is the basis for testing H2a and H2b, provides data on machine interruptions that were solved by workers. We excluded production days with reduced working hours and weekends, as for the production efficiency data set.

Identification Strategy

In order to examine the impact of the *choice* system on production and task performance, we utilize a DiD estimation strategy. The implementation of the *choice* system is the treatment that the treatment group received while the control group continued to operate as before. The treatment and control group belong to different plants of the same company, which operates with standardized processes, routines, and management styles across all locations. All company-wide changes that are exogenous to our model affect both groups simultaneously. Furthermore, both groups have worked with the initial *delegation* version of the smartwatch-based technology for at least six months prior to the treatment. Thus, both groups share comparable levels of competence with the smartwatch-based technology. The treatment and control groups perform the same process step (i.e., molding), use similar machinery (i.e., pressing machines), and produce similarly complex parts (e.g., small gearwheels). Table1 provides an overview of the groups. We verify the comparability further in several robustness checks.

Table 1: Overview of treatment and control group

Characteristic	Treatment Group	Control Group
Number of workers	56	46
Number of machines	53	50
Number of tasks per worker[a]	620	811
Avg. Machine Interruptions[b]	0,30	0,24

[a]Unique tasks during time of field experiment (09. Jul. 2021 – 10. Oct. 2021)
[b]per hour and machine.

DiD is a well-established econometric approach, for instance, for assessing the effect of policies (Card & Krueger, 2000), and has successfully been applied to operations management-related issues (Lu & Lu, 2017; Pierce et al., 2015; Staats et al., 2017; Tan & Netessine, 2020). We employ the subsequent model structure to evaluate the impact of the *choice* system on production efficiency, and task efficiency, and responsiveness:

$$Y_{it} = \alpha_0 + \beta_1 \times TreatedGroup_i + \beta_2 \times IsPost_t + \beta_3 \times TREAT_{it} + \beta_4 \times MorningShift_t + \beta_5 \times AfternoonShift_t + \beta_6 \times Utilization_{it} + \epsilon_{it} \quad (1)$$

where Y_{it} is the dependent variable for group i at hour t, α is the model intercept, $TreatedGroup_i$ is a dummy variable that identifies the treatment group, $IsPost_i$ is a dummy variable that identifies the post-treatment period (0 before and 1 after the treatment), $TREAT_{it}$ is the interaction of $TreatedGroup_i$ and $IsPost_i$. This interaction effect is of the strongest interest in our analyses. It denotes the difference over time in the difference among the groups (DiD). Additionally, $MorningShift_t$ and $AfternoonShift_t$ are dummy variables that identify all three shifts and $Utilization_{it}$ captures possible influences of varying utilization. All of our dependent variables are continuous variables at the group level. We use ordinary least squares (OLS) regressions in all estimations.

RESULTS

Treatment Effects

Table 2 presents the treatment effects of the *choice* system on the treatment group. We provide the full regression output of the models in Appendix A.

Table 2: Impact of Autonomy on Production and Worker Performance

Dependent Variable:	(1)[a] Production efficiency	(2)[b] Task efficiency	(3)[b] Task responsiveness
Treated with *choice*	0.069***	-17.496	-73.215***
Time fixed effects	Included	Included	Included
Group fixed effects	Included	Included	Included
Constant	0.916***	9685.981***	2568.324***
Adjusted R^2	0.093	0.003	0.055
Observations	2926	2901	2901

Notes: Regressions use robust standard errors. [a]data from machines/MES system [b]data from smartwatch system ***Significant at the .05% confidence level

The models contain group fixed effects, which account for a time invariant difference between treatment and control group and fixed effects that control for trends. Column (1) shows that the introduction of autonomy increased production efficiency by 6.9% with a 95% interval that ranges from 4.5% to 9.3%. This represents an average increase of 57 pieces per hour or 1,368 pieces per day. Considering the average price per piece of USD 0,191 [This is the average price per part after the process step of molding. The sales price is higher.], this accounts for an average increase in revenue of ca. USD 64,000 per year. Column (2) shows that there is no significant impact of *choice* on task efficiency. Column (3) indicates that autonomy impacts task responsiveness negatively. Tasks wait, on average, 73 seconds longer until acceptance by a worker when compared to the *delegation* system. It has a 95% interval between -98.8 seconds and -47.6 seconds. In addition, we provide descriptive statistics in Appendix B.

Robustness Checks

Parallel Trends Assumption

The DiD method's credibility depends on an essential premise, known as parallel trends, which asserts that the treated and control groups would have had comparable performance trends in the absence of the treatment. We support this assumption with detailed knowledge of the two groups such as their production process (see section "Identification Strategy") and by visually examining the graphical trends of the dependent variables. In addition, we show that the trends of both groups are not significantly different by conducting statistical tests on the average growth rates of the dependent variables during the pre-treatment period.

We use three graphs to demonstrate the parallel trends for the treated and control group in hourly production efficiency rates as well as hourly task efficiency and responsiveness rates before the treatment. Our data set captures these rates for a time-period of nine weeks before the treatment (09. Jul. 2021–17. Sept. 2021), the longest possible time-period without a major system change. Figure 2-4 show the hourly moving average of production efficiency, task efficiency, and responsiveness of both the treated and the control groups before the treatment. Both the treated group and the control group appear to have similar linear trends when fitted. Also, the hourly data shows clear parallel developments. Note that performance differences between the two groups (i.e., the distance between the linear trends) are captured by TreatedGroupi in our estimations and therefore do not affect the hypothesis tests.

Figure 2: Ex-ante Parallel Trend for Production Efficiency

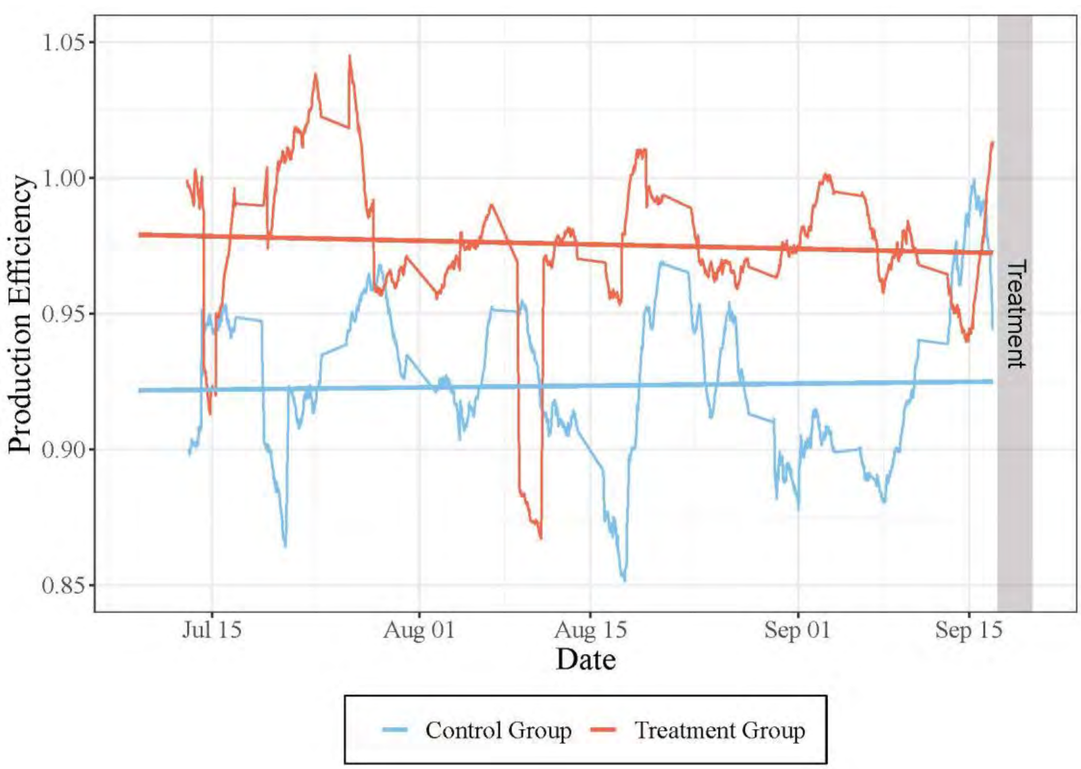
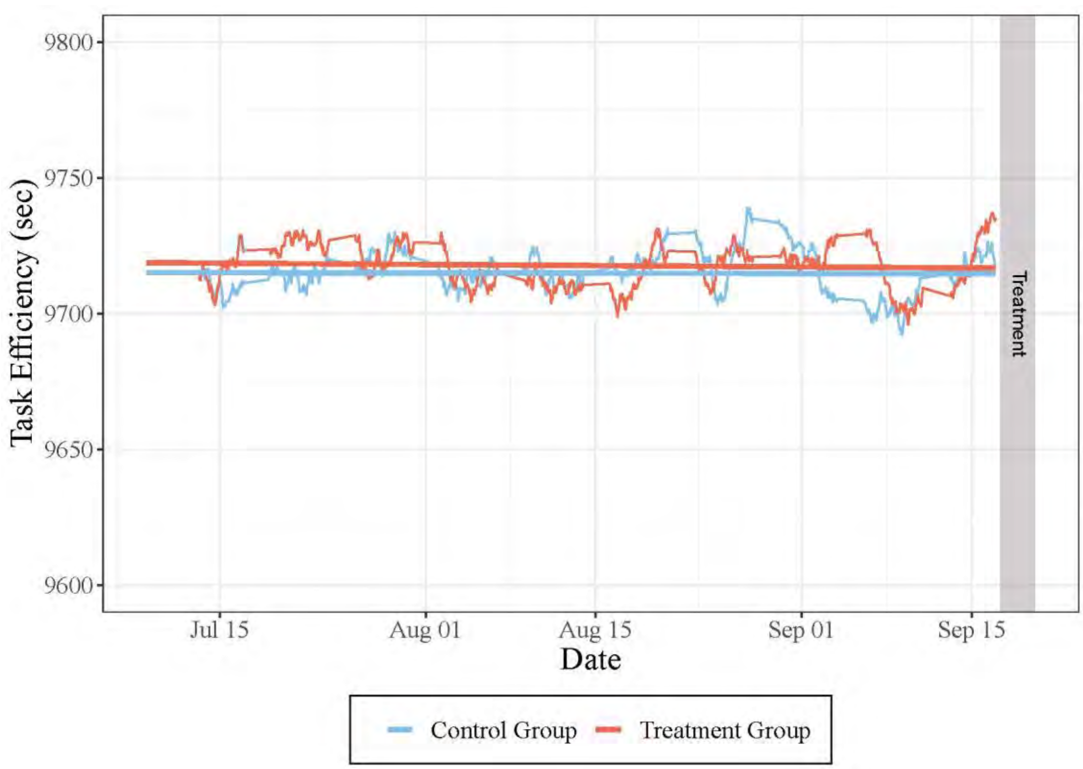


Figure 3: Ex-ante Parallel Trend for Task Responsiveness

Figure 4: Ex-ante Parallel Trend for Task Efficiency



We analytically validate the parallel trends assumption by testing whether the average relative performance difference between treatment and control group changed over time prior to the treatment. Therefore, we estimate an interaction term between the performance difference and a continuous variable that increments for each consecutive shift. A significant interaction would indicate that the control and treatment groups changed their relative performance even before our treatment, making the pair infeasible for further analysis. The results suggest that the relative difference of both groups did not significantly change over time before the treatment for production efficiency (p = .972), task efficiency (p = .072), and task responsiveness (p = .710). We repeat the estimation for a modified time-period that is six weeks long instead of nine weeks (02. Aug. 2021–17. Sept. 2021). Also, the modified time-period supports the parallel trend assumption. We provide the results in Appendix C.

<u>Hawthorne Effect</u>

The Hawthorne effect is a possible confounding factor that could distort the actual impact of the treatment on worker performance. It can create a temporary shift in performance due to workers' awareness of being monitored (Landsberger, 1957). We minimized the potential Hawthorne effect by asking the company management to conduct "business as usual" to avoid unusual management attention on the shop floor. Only team leaders, who always work on the shop floor, represented management on the shop floor during the study. Furthermore, the research team minimized visits during the time of the field experiment to the day of the implementation to answer questions. No further visits happened during the data collection to reduce possible Hawthorne effects.

Management announced the system change one week prior to its implementation in the treatment group, as all changes are communicated ahead of time. If a Hawthorne effect affected our results, it would have shown from that day on. To further inspect this, we assess production efficiency in the two weeks before the treatment and separate the first one (before announcement) from the second one (after the announcement). We find no significant difference using the DiD estimation technique (i.e., p = .07) between the treatment and control group. Furthermore, if workers had anticipated evaluation, they would have attempted to improve all performance dimensions. We only find improvements in one of three variables.

<u>Placebo Test</u>

We mitigate the possibility of identifying false-positive outcomes in our investigation (Bertrand et al., 2004). We run a placebo test that randomly re-allocates the treatment to 60 different periods as suggested in (Pierce et al., 2015). We run the test for the significant results found for production efficiency and task responsiveness. Figure 5 and 6 show the results of the placebo tests for production efficiency and task responsiveness, respectively. Each point represents the estimate obtained for the treatment along with its 95% confidence interval. The estimates are ranked in size and the estimate of the true treatment is the right-most estimate. In the production efficiency model (Figure 5), the placebo point estimates range between -0.032 and 0.038, while the actual estimate is close to 0.05—approximately 24% higher than the largest placebo point estimate. Similarly, in the task responsiveness model (Figure 6), the placebo point estimates range between 25 and -5 seconds , whereas the actual estimate is close to -50 seconds—10 times lower than the lowest placebo point estimate. The placebo tests indicate that our results are not false positives.

Figure 5: Placebo tests for production efficiency

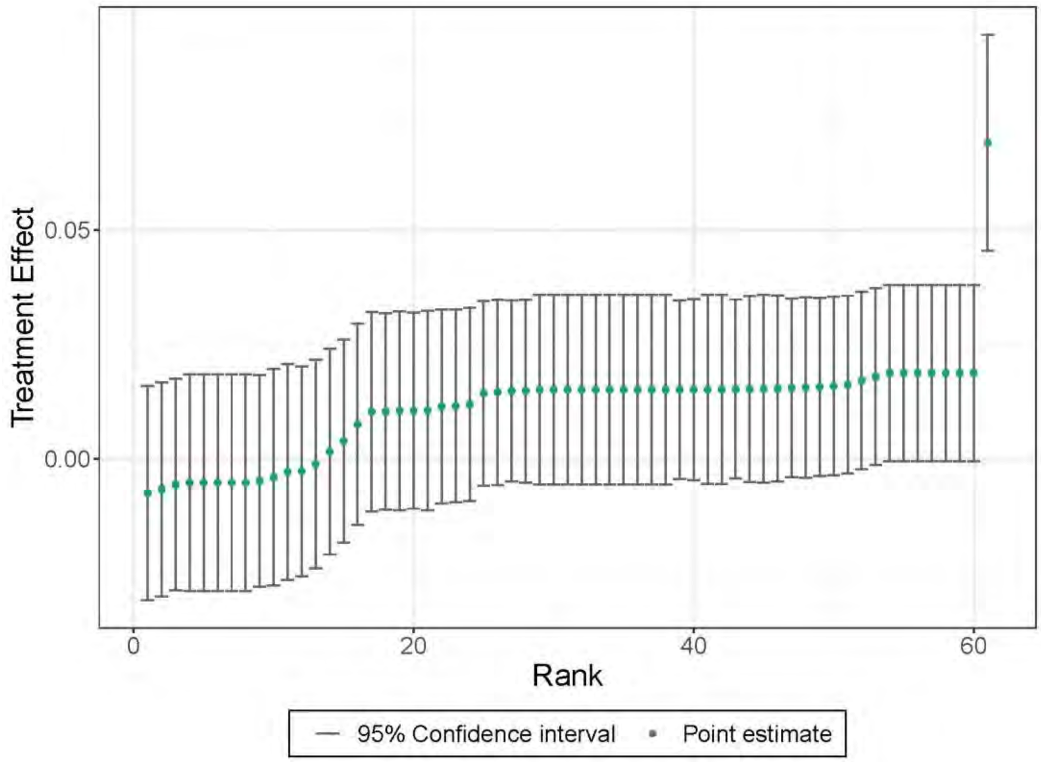
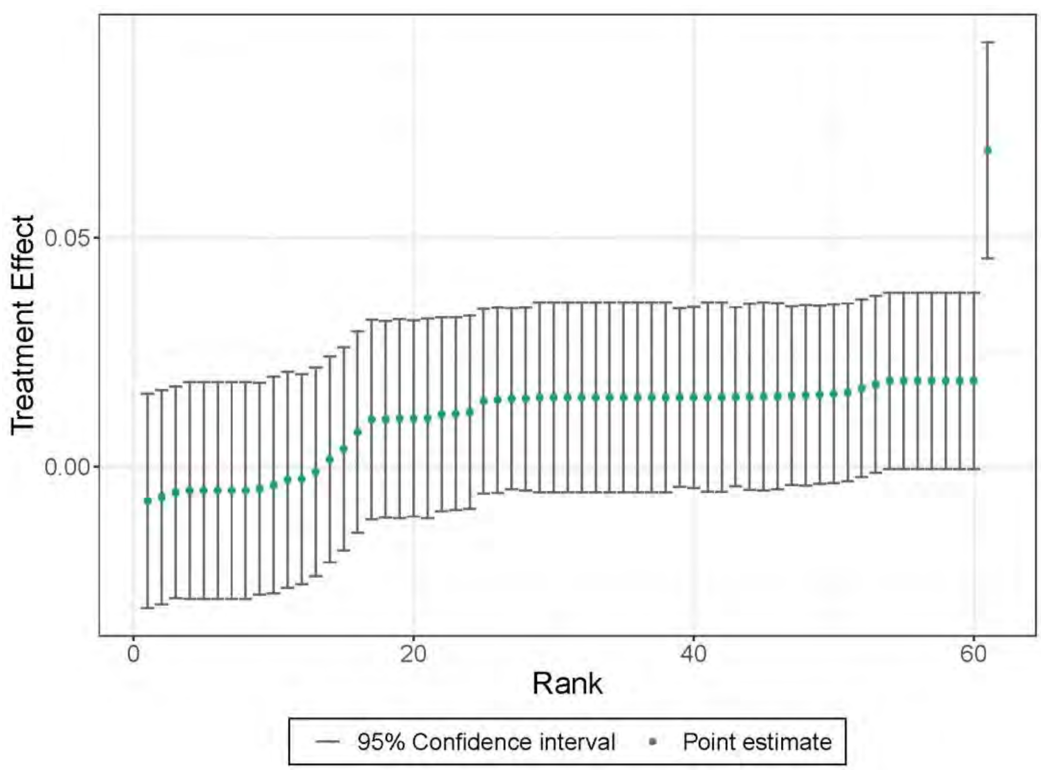


Figure 6: Placebo tests for production efficiency



Machine-Level and Task-Level Analysis

In addition to the group-level analysis, we conduct additional analyses on the machine level for production efficiency and task level for task efficiency and responsiveness. These disaggregated analyses provide robustness to the main analyses on the group level. For machine-level analysis of production efficiency, we use machine IDs to distinguish between machines from the treatment and control group. For task efficiency and responsiveness, we use task IDs to identify whether a task has been completed in the treatment or control group. Specifically, we employ the following machine-level and task-level analysis:

$$Y_{ijt} = \alpha_0 + \beta_1 \times TreatedMachine_{ij} + \beta_2 \times IsPost_t + \beta_3 \times TREAT_{it} + \beta_4 \times MorningShift_t + \beta_5 \times AfternoonShift_t + \beta_6 \times Utilization_{it} + \epsilon_{it} \quad (2)$$

$$Y_{ijt} = \alpha_0 + \beta_1 \times TreatedTask_{ij} + \beta_2 \times IsPost_t + \beta_3 \times TREAT_{ijt} + \beta_4 \times MorningShift_t + \beta_5 \times AfternoonShift_t + \beta_6 \times Utilization_{ijt} + \epsilon_{ijt} \quad (3)$$

where Y_{ijt} is the dependent variable for machine j (or task j) at group i at datetime t, $TreatedMachine_{ij}$ (or $TreatedTask_{ij}$) is a dummy variable that identifies machines/tasks exposed to the treatment, $TREAT_{ijt}$ is the interaction of $TreatedMachine_{ij}$ (or $TreatedTask_{ij}$) and $IsPost_i$, and $Utilization_{ijt}$ is the utilization of machine j at group i at datetime t. The remaining variables remain unchanged. Table 3 provides the treatment effect for the disaggregated analysis.

Table 3: Machine- and Task-Level Analysis of Impact of Autonomy on Production and Worker

	(1)	(2)	(3)
Dependent Variable:	Production efficiency	Task efficiency	Task responsiveness
Aggregation Level:	Machine-level	Task-level	Task-level
Treated with *choice*	0.038***	12.670	-73.953***
Time fixed effects	Included	Included	Included
Group fixed effects	Included	Included	Included
Constant	0.909***	9690.195***	18283.787***
Adjusted R^2	0.016	0.002	0.012
Observations	65,348	24,725	24,725

Notes: Regressions use robust standard errors. ***Significant at the .05% confidence level

The machine-level analysis mirrors our main results for production efficiency. As with the group-level results, Column (1) shows that autonomy increases production efficiency by 3.8%. It has a 99% confidence interval that ranges between 2.4% and 5.2%. The task-level analysis also mirrors our main results for task efficiency and responsiveness. Column (2) indicates that the establishment of autonomy has no impact on task efficiency on the task-level. Column (3) shows that autonomy decreased task responsiveness by 73 seconds. Its 99% confidence interval is between -94.226 and -53.680 seconds. We repeat this disaggregated analysis for our modified time-period and again find consistency in the results. We provide the results for the modified time-periods in Appendix D.

Shift-Level Analysis

In addition to the disaggregated analysis of structural features (i.e., machines or tasks), we conduct an analysis at the shift level, thus aggregating the hourly data of the main analysis. Specifically, we employ the following shift-level model:

$$Y_{ijt} = \alpha_0 + \beta_1 \times TreatedShift_{ij} + \beta_2 \times IsPost_t + \beta_3 \times TREAT_{ijt} + \beta_4 \times MorningShift_t + \beta_5 \times AfternoonShift_t + \beta_6 \times Utilization_{ijt} + \epsilon_{ijt} \quad (4)$$

where Y_{ijt} is the dependent variable for shift j at group i at date t, $TreatedShift_{ij}$ is a dummy variable that identifies the shifts exposed to the treatment, and $Utilization_{ijt}$ is the utilization of shift j at group i at date t. The remaining variables remain unchanged. Table 4 shows the treatment effect for the shift-level analysis.

Table 4: Shift-Level Analysis of Impact of Autonomy on Production and Worker Performance

	(1)	(2)	(3)
Dependent Variable:	Production efficiency	Task efficiency	Task responsiveness
Aggregation Level:	Shift-level	Shift-level	Shift-level
Treated with *choice*	0.071***	12.267	-65.066***
Time fixed effects	Included	Included	Included
Group fixed effects	Included	Included	Included
Constant	0.914***	865.881***	491.772***
Adjusted R^2	0.169	0.083	0.202
Observations	382	401	401

Notes. Regressions use robust standard errors. ***Significant at the .05% confidence level

Again, the shift-level analysis mirrors the conclusions drawn from the earlier analyses . Column (1) shows that *choice* increases production efficiency by 7.1% (99% confidence interval ranges between 2.5% and 11.8%). Column (2) indicates that *choice* has no impact on task efficiency. Column (3) reveals that *choice* decreases task responsiveness by 65.1 seconds (99% confidence interval ranges from -95.87 seconds to -34.26 seconds). We repeat our shift-level analysis for our modified time-period and find consistency in our results. We provide the results for the modified time-periods in Appendix E.

DISCUSSION

Our field experiment assesses a change in a fully digitalized production system that makes use of a state-of-the-art internet-of-things technology based on smartwatches. The use of smartwatches enabled the company to fundamentally transform the shop floor organization from a traditional organization that assigns workers to machines into a much more flexible organization of work based on tasks, regardless of where they occur. Few scholars have scrutinized such fine-grained organization of work in operations and supply chain management (e.g., Bals & Turkulainen, 2017; Hoogeweegen et al., 1999), and no study has, to the best of our knowledge, examined a fully digitalized production environment based on tasks or observed changes in such systems. Our research studies a much more granular task-based organization compared to earlier research that segmented tasks in purchasing processes (Bals & Turkulainen, 2017).

Our results suggest that not only the technical features of a cyber-physical production system are important but also that human sensemaking plays an essential role in production. Particularly, increasing autonomy in such systems can lead to positive and negative changes in performance. Our study shows that providing autonomy by means of digital technology generates benefits in overall production efficiency. This is an encouraging sign for digitalization in manufacturing and parallels earlier evidence that non-interdependent tasks, such as in our study, benefit from increasing individual autonomy (Langfred, 2005). This finding should be probed in interdependent manufacturing contexts, such as assembly lines, in the future. Our results are also consistent with studies that suggest autonomy as a performance-enhancing variable in unpredictable task environments (Cordery et al., 2010), such as ours where interruptions occur spontaneously. Thus, future research should examine our results in digitalized production facilities where tasks appear according to a predetermined schedule, for example, and can also examine the effects of task variety vis-à-vis our study that focused on machine interrupts alone (e.g., Dodd & Ganster, 1996).

One of the unexpected results of our study is related to task efficiency, which did not improve after increasing autonomy. One possible explanation in our case may be the factory's utilization, which was operating at its capacity limits. Even before the system change, "37% of all machine interruptions could not be solved in time," according to the shop floor digitalization project leader of the treatment group. Considering earlier studies on workload effects on efficiency (e.g., Kc & Terwiesch, 2009; Oliva & Sterman, 2001), our study provides evidence that increased autonomy does not increase task efficiency when workers are operating at peak capacity. This is in line with Kc et al. (2020), who showed that choosing easier tasks at high workloads can even have negative performance implications. Thus, a set of machines and operators with a lower workload may be more likely to show efficiency gains. This finding indicates that the utilization of a factory may be a boundary condition for self-determination theory, as applied in our study.

Another explanation for the absence of an effect on task efficiency may relate to the degree to which our host company fulfilled the job design criteria for autonomous work (Hackman & Lorsch, 1987; Wall et al., 1986). For instance, Wall et al. (1986) accredited a null effect to incorrectly distributed rewards and unclear performance feedback. However, workers did not receive rewards and we did not provide performance feedback. However, the job itself provides an explanation. Involving workers in the entire work process instead of only fragmented parts and showing workers that their job affects the well-being of others are two facets that Hackman and Oldham (1975) stresses as important conditions for autonomy to show an effect . In our study, we can assume that both task characteristics were low since workers were only responsible for a relatively small part of the work (i.e., molding) and since resolving machine interruptions had no visible effect on their co-workers. Thus, the lack of interwovenness of job characteristics and autonomy, as put forward in the job characteristics approach, may explain the absence of an effect of autonomy on task efficiency. At the same time, it points to potential future laboratory work that manipulates autonomy, as our study did, and also manipulates job characteristics put forward by the job characteristics model (Hackman & Oldham, 1975).

Our work shows that providing autonomy in digital production systems promises gains but also risks slower task responsiveness. In the *delegation* scenario, lower autonomy seems to have urged workers to respond more quickly to tasks that are pushed to their devices, as they cannot see and choose other open tasks in the backlog (see Figure 1). Workers may have responded to the relatively aggressive presentation of work tasks in the *delegation* system and to the implicit management expectation to accept. Conversely, more autonomy gave workers the freedom to make their own choices, which takes time and may have contributed to the longer response times. This finding implies that increased autonomy in operations can also have negative impacts on task performance features.

This augments the pervasive view of self-determination theory that increased autonomy generally leads to an improvement in performance outcomes (Deci & Ryan, 1980, 1985; Ryan & Deci, 2000). We add that increases in information processing needs that can slow the reaction times of production workers. The magnitude of the processing need could be an explanation for unresolved conflicts in the literature around autonomous choices and responsiveness: laboratory studies found that relatively simple sequencing choices improve reaction times (Ziv & Lidor, 2021) whereas a choice among alternative incentive schemes has no effect (Williams & Luthans, 1992). Compared to these laboratory studies, the information-processing need of a choice is relatively high in our real-world field experiment. This indicates a possible hidden moderator and cautions management, as well as future scholarship, to also consider the information richness of an autonomous choice. This is particularly relevant for production and operations management, since digitalization strongly drives the density and richness of operational information in manufacturing but also seeks to empower shop floor employees at the same time (Leyer et al., 2019). Thus, our study should motivate research to further examine the link between the complexity of autonomous choices, particularly in operational contexts that rely even more on quick and accurate responses, such as in healthcare or disaster relief.

CONCLUSION

In this study, we analyze 65,348 machine status reports and 49,050 work tasks gathered in a digitalized production system based on smartwatch technology. We estimate the causal effect of providing more autonomy in the task allocation process using a DiD approach. Our results suggest that providing autonomy in task allocation in a digitalized production system hold ambivalent implications for performance outcomes. We find that production efficiency increases and task responsiveness decreases at the same time, when workers are able to autonomously choose tasks from a list of pending production tasks instead of being assigned a task from the system (see Figure 1). Even though workers responded 73 seconds slower to tasks, on average, the efficiency in the molding process increased by 6.9% — that is an increase of 1,368 parts per day and ca. USD 65,000 p.a. Thus, our findings encourage managers to grant more autonomy in their digital production systems. However, they also remind managers to closely examine how increasing autonomy affects the complexity of workers' decision making and whether possible delays can disturb the production process.

Our study has several limitations. First, the plants were running close to capacity limits. In addition, personnel shortages due to the ongoing COVID-19 pandemic and full order books have led to a burdensome work environment. This might have affected our field experiment. To increase the robustness of our results, the field experiment could be replicated in a plant with more slack or in future laboratory research. This research may further explore the exact cut-off point where autonomy stop contributing to efficiency as utilization increases. A second limitation we identify refers to the task characteristics of the conducted field experiment. Specifically, workers had control over only a small part of the value chain and no visible effect on others' work, which may be an explanation for the null results of autonomy (Wall et al. 1986). Thus, we call for more research to conduct similar experiments, including tasks with different job characteristics, to examine if our results hold. Lastly, our study is limited to one specific digital production technology and batch-type production. We encourage future research to conduct similar studies with other digital production technology or other production types, such as classical assembly lines or job shop production.

In our study, the *choice* system showed an important improvement compared to the prior *delegation* system in terms of production efficiency. Yet, the scholarly work to find the best performing and most socially acceptable digital production systems is not over but has just begun.

REFERENCES

Arregle, J.-L., Dattée, B., Hitt, M. A., & Bergh, D. (2023). Organizational Autonomy: A Review and Agenda for Future Research. *Journal of Management*, *49*(1), 85-124.

Ashkanani, A. M., Dunford, B. B., & Mumford, K. J. (2022). Impact of Motivation and Workload on Service Time Components: An Empirical Analysis of Call Center Operations. *Management Science*, *68*(9), 6697-6715.

Bachrach, D. G., & Bendoly, E. (2011). Rigor in behavioral experiments: A basic primer for supply chain management researchers. *Journal of Supply Chain Management*, *47*(3), 5-8.
Baldwin, C. Y., & Clark, K. B. (2006). The architecture of participation: Does code architecture mitigate free riding in the open source development model? *Management Science*, *52*(7), 1116-1127.
Bals, L., & Turkulainen, V. (2017). Achieving efficiency and effectiveness in Purchasing and Supply Management: Organization design and outsourcing. *Journal of Purchasing and Supply Management*, *23*(4), 256-267.
Barrick, M. R., & Mount, M. K. (1993). Autonomy as a moderator of the relationships between the Big Five personality dimensions and job performance. *Journal of applied psychology*, *78*(1), 111-118.
Basu, S., & Savani, K. (2017). Choosing one at a time? Presenting options simultaneously helps people make more optimal decisions than presenting options sequentially. *Organizational Behavior and Human Decision Processes*, *139*, 76-91.
Baumers, M., & Holweg, M. (2019). On the economics of additive manufacturing: Experimental findings. *Journal of Operations Management*, *65*(8), 794-809.
Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249-275.
Bledow, R., Kühnel, J., Jin, M., & Kuhl, J. (2022). Breaking the chains: The inverted-U-shaped relationship between action-state orientation and creativity under low job autonomy. *Journal of Management*, *48*(4), 905-935.
Boudreau, J., Hopp, W., McClain, J. O., & Thomas, L. J. (2003). On the interface between operations and human resources management. *Manufacturing & Service Operations Management*, *5*(3), 179-202.
Breaugh, J. A. (1985). The Measurement of Work Autonomy. *Human Relations*, *38*(6), 551-570.
Cagliano, R., Canterino, F., Longoni, A., & Bartezzaghi, E. (2019). The interplay between smart manufacturing technologies and work organization: the role of technological complexity. *International Journal of Operations & Production Management*, *39*(8), 913-934.
Card, D., & Krueger, A. B. (2000). Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania: reply. *American Economic Review*, *90*(5), 1397-1420.
Cordery, J. L., Morrison, D., Wright, B. M., & Wall, T. D. (2010). The impact of autonomy and task uncertainty on team performance: A longitudinal field study. *Journal of Organizational Behavior*, *31*(2 - 3), 240-258.
De Quidt, J., Haushofer, J., & Roth, C. (2018). Measuring and bounding experimenter demand. *American Economic Review*, *108*(11), 3266-3302.
Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of applied psychology*, *74*(4), 580-590.
Deci, E. L., Olafsen, A. H., & Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual review of organizational psychology and organizational behavior*, *4*, 19-43.
Deci, E. L., & Ryan, R. M. (1980). Self-determination theory: When mind mediates behavior. *The Journal of mind and Behavior*, *1*(1), 33-43.
Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109-134.
Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227-268.
Dodd, N. G., & Ganster, D. C. (1996). The interactive effects of variety, autonomy, and feedback on attitudes and performance. *Journal of Organizational Behavior*, *17*(4), 329-347.

Donald, J. N., Bradshaw, E. L., Conigrave, J. H., Parker, P. D., Byatt, L. L., Noetel, M., & Ryan, R. M. (2021). Paths to the light and dark sides of human nature: A meta-analytic review of the prosocial benefits of autonomy and the antisocial costs of control. *Psychological bulletin*, *147*(9), 921-946.

Donohue, K., Özer, Ö., & Zheng, Y. (2020). Behavioral operations: Past, present, and future. *Manufacturing & Service Operations Management*, *22*(1), 191-202.

Eckerd, S., DuHadway, S., Bendoly, E., Carter, C. R., & Kaufmann, L. (2021). On making experimental design choices: Discussions on the use and challenges of demand effects, incentives, deception, samples, and vignettes. *Journal of Operations Management*, *67*(2), 261-275.

Eisenberger, R., Armeli, S., Rexwinkel, B., Lynch, P. D., & Rhoades, L. (2001). Reciprocation of perceived organizational support. *Journal of applied psychology*, *86*(1), 42-51.

Ernst, A., Jiang, H., & Krishnamoorthy, M. (2006). Exact solutions to task allocation problems. *Management Science*, *52*(10), 1634-1646.

Ewing, J., & Cohen, P. (2021). How Car Shortages Are Putting the World's Economy at Risk. *The New York Times, Article L*, *1360*.

Grant, A. M., Nurmohamed, S., Ashford, S. J., & Dekas, K. (2011). The performance implications of ambivalent initiative: The interplay of autonomous and controlled motivations. *Organizational Behavior and Human Decision Processes*, *116*(2), 241-251.

Hackman, J. R., & Lorsch, J. (1987). Handbook of organizational behavior. *Handbook of Organizational Behavior, Prentice-Hall, Englewood Cliffs, NJ*, 315-342.

Hackman, J. R., & Oldham, G. R. (1975). Development of the job diagnostic survey. *Journal of applied psychology*, *60*(2), 159-170.

Heinen, J. J., & Hoberg, K. (2019). Assessing the potential of additive manufacturing for the provision of spare parts. *Journal of Operations Management*, *65*(8), 810-826.

Holmström, J., Holweg, M., Lawson, B., Pil, F. K., & Wagner, S. M. (2019). The digitalization of operations and supply chain management: Theoretical and methodological implications. In (Vol. 65, pp. 728-734): Wiley Online Library.

Hoogeweegen, M. R., Teunissen, W. J., Vervest, P. H., & Wagenaar, R. W. (1999). Modular network design: Using information and communication technology to allocate production tasks in a virtual organization. *Decision Sciences*, *30*(4), 1073-1103.

Işik, T., Andradóttir, S., & Ayhan, H. (2021). Dynamic Control of Non‐Collaborative Workers When Reassignment Is Costly. *Production and Operations Management*, *31*(3), 1332-1352.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, *79*(6), 995-1006.

Jean, R. J. B., Sinkovics, R. R., & Hiebaum, T. P. (2014). The Effects of Supplier Involvement and Knowledge Protection on Product Innovation in Customer–Supplier Relationships: A Study of Global Automotive Suppliers in C hina. *Journal of Product Innovation Management*, *31*(1), 98-113.

Jordan, W. C., & Graves, S. C. (1995). Principles on the benefits of manufacturing process flexibility. *Management Science*, *41*(4), 577-594.

Kamei, K., & Markussen, T. (2022). Free Riding and Workplace Democracy—Heterogeneous Task Preferences and Sorting. *Management Science*, 1-21.

Kc, D. S., Staats, B. R., Kouchaki, M., & Gino, F. (2020). Task selection and workload: A focus on completing easy tasks hurts performance. *Management Science*, *66*(10), 4397-4416.

Kc, D. S., & Terwiesch, C. (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, *55*(9), 1486-1498.

Kembro, J. H., Danielsson, V., & Smajli, G. (2017). Network video technology: Exploring an innovative approach to improving warehouse operations. *International Journal of Physical Distribution & Logistics Management*, *47*(7), 623-645.

Kim, T. Y., Cable, D. M., Kim, S. P., & Wang, J. (2009). Emotional competence and work performance: The mediating effect of proactivity and the moderating effect of job autonomy. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *30*(7), 983-1000.

Landsberger, H. A. (1957). *Hawthorne revisited: A plea for an open city*. Cornell University.

Langfred, C. W. (2005). Autonomy and performance in teams: The multilevel moderating effect of task interdependence. *Journal of Management*, *31*(4), 513-529.

Langfred, C. W., & Moye, N. A. (2004). Effects of task autonomy on performance: an extended model considering motivational, informational, and structural mechanisms. *Journal of applied psychology*, *89*(6), 934-945.

Lewis, G. H., Srinivasan, A., & Subrahmanian, E. (1998). Staffing and allocation of workers in an administrative office. *Management Science*, *44*(4), 548-570.

Leyer, M., Richter, A., & Steinhüser, M. (2019). “Power to the workers”: Empowering shop floor workers with worker-centric digital designs. *International Journal of Operations & Production Management*, *39*(1), 24-42.

Lim, Y. F., & Wu, Y. (2014). Cellular bucket brigades on U‐lines with discrete work stations. *Production and Operations Management*, *23*(7), 1113-1128.

Lu, S. F., & Lu, L. X. (2017). Do mandatory overtime laws improve quality? Staffing decisions and operational flexibility of nursing homes. *Management Science*, *63*(11), 3566-3585.

Maghazei, O., Lewis, M. A., & Netland, T. H. (2022). Emerging technologies and the use case: A multi‐year study of drone adoption. *Journal of Operations Management*, *68*(6-7), 560-591.

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research—Understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research*, *44*(2), 465-476.

Morgeson, F. P., Delaney-Klinger, K., & Hemingway, M. A. (2005). The importance of job autonomy, cognitive ability, and job-related skill for predicting role breadth and job performance. *Journal of applied psychology*, *90*(2), 399-406.

OECD. (2017). *The Next Production Revolution*. Retrieved 05/31/2023 from http://dx.doi.org/https://doi.org/https://doi.org/10.1787/9789264271036-en

Oliva, R., & Sterman, J. D. (2001). Cutting corners and working overtime: Quality erosion in the service industry. *Management Science*, *47*(7), 894-914.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychological bulletin*, *134*(2), 270-300.

Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science*, *61*(10), 2299-2319.

Raveendran, M., Puranam, P., & Warglien, M. (2021). Division of labor through self-selection. *Organization Science*, *33*(2), 1-21.

Rea, D., Froehle, C., Masterson, S., Stettler, B., Fermann, G., & Pancioli, A. (2021). Unequal but Fair: Incorporating Distributive Justice in Operational Allocation Models. *Production and Operations Management*, *30*(7), 2304-2320.

Roehrich, J. K., Hoejmose, S. U., & Overland, V. (2017). Driving green supply chain management performance through supplier selection and value internalisation: A self-determination theory perspective. *International Journal of Operations & Production Management*, *37*(4), 489-509.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68-78.

Sarafan, M., Lawson, B., Roehrich, J. K., & Squire, B. (2022). Knowledge sharing in project-based supply networks. *International Journal of Operations & Production Management*, *42*(6), 852-874.

Senoner, J., Netland, T., & Feuerriegel, S. (2022). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science*, *68*(8), 5704-5723.

Sharma, L., Queenan, C., & Ozturk, O. (2019). The Impact of Information Technology and Communication on Medical Malpractice Lawsuits. *Production and Operations Management*, *28*(10), 2552-2572.

Shunko, M., Niederhoff, J., & Rosokha, Y. (2018). Humans are not machines: The behavioral impact of queueing design on service time. *Management Science*, *64*(1), 453-473.

Spreitzer, G., Sutcliffe, K., Dutton, J., Sonenshein, S., & Grant, A. M. (2005). A socially embedded model of thriving at work. *Organization Science*, *16*(5), 537-549.

Staats, B. R., Dai, H., Hofmann, D., & Milkman, K. L. (2017). Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Science*, *63*(5), 1563-1585.

Ta, H., Esper, T. L., & Tokar, T. (2021). Appealing to the Crowd: Motivation Message Framing and Crowdsourcing Performance in Retail Operations. *Production and Operations Management*, *30*(9), 3192-3212.

Tan, T. F., & Netessine, S. (2020). At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science*, *66*(10), 4496-4515.

Van den Broeck, A., Ferris, D. L., Chang, C.-H., & Rosen, C. C. (2016). A review of self-determination theory's basic psychological needs at work. *Journal of Management*, *42*(5), 1195-1229.

Wall, T. D., Kemp, N. J., Jackson, P. R., & Clegg, C. W. (1986). Outcomes of autonomous workgroups: A long-term field experiment. *Academy of Management journal*, *29*(2), 280-304.

Williams, S., & Luthans, F. (1992). The impact of choice of rewards and feedback on task performance. *Journal of Organizational Behavior*, *13*(7), 653-666.

World Bank. (2023). *Manufacturing, value added (% of GDP)*. Retrieved 31/5/2023 from https://data.worldbank.org/indicator/NV.IND.MANF.ZS?most_recent_value_desc=true

Zavadlav, E., McClain, J. O., & Thomas, L. J. (1996). Self-buffering, self-balancing, self-flushing production lines. *Management Science*, *42*(8), 1151-1164.

Ziv, G., & Lidor, R. (2021). Autonomy support and preference-performance dissociation in choice-reaction time tasks. *Human Movement Science*, *77*, 102786.

APPENDIX A: DID REGRESSION OUTPUT

Table 5: Production Efficiency DiD Regression Output

Coefficients	Estimates	Std. Error	t-Statistic	p-value
Intercept	0.916	.006	165.428	.000
Treated_Group	0.052	.005	10.254	.000
Is_Post	-0.008	.012	-0.676	.500
Morning_Shift	0.006	.006	0.934	.350
Afternoon_Shift	0.017	.005	3.333	.001
Utilization	-0.010	.046	-0.221	.083
Treated_Group*Is_Post	0.069	.013	5.193	.000

Note: Regression uses robust standard errors.

Table 6: Task Efficiency DiD Regression Output

Coefficients	Estimates	Std. Error	t-Statistic	p-value
Intercept	9685.981	39.746	243.698	.000
Treated_Group	27.966	6.736	4.152	.000
Is_Post	-2.983	14.606	-0.204	.838
Morning_Shift	15.024	7.658	1.962	.050
Afternoon_Shift	-6.199	11.968	-0.518	.605
Utilization	-17.385	55.326	-0.314	.753
Treated_Group*Is_Post	-17.496	28.678	-0.610	.000

Note: Regression uses robust standard errors.

Table 7: Task Responsiveness DiD Regression Output

Coefficients	Estimates	Std. Error	t-Statistic	p-value
Intercept	2568.324	19.247	133.439	.000
Treated_Group	65.674	4.722	13.908	.000
Is_Post	43.296	9.658	4.483	.000
Morning_Shift	-24.761	6.564	-3.772	.000
Afternoon_Shift	-3.615	6.105	-0.592	.534
Utilization	17.393	26.087	0.667	.505
Treated_Group*Is_Post	-73.215	17.669	-4.144	.000

Note: Regression uses robust standard errors.

APPENDIX B: DESCRIPTIVE STATISTICS

Table 8: Descriptive Statistics for Production Efficiency

	Scenario	N	Mean	Std. Deviation	Std. Error Mean
Production Efficiency	*delegation*	31,834	0.962	0.232	0.001
	choice	7,575	1.015	0.295	0.003

Note: N sub-sample size on hourly basis; data stems directly from the machines.

Table 9: Descriptive Statistics for Task Responsiveness and Efficiency

		delegation		*choice*	
	N	Mean	Std. Dev.	Mean	Std. Dev
Task Responsiveness[a]	1,509	136.3 secs	83.2	167.1 secs	230.0
Task Efficiency[a]	1,509	148.2 secs	89.6	167.0 secs	541.2

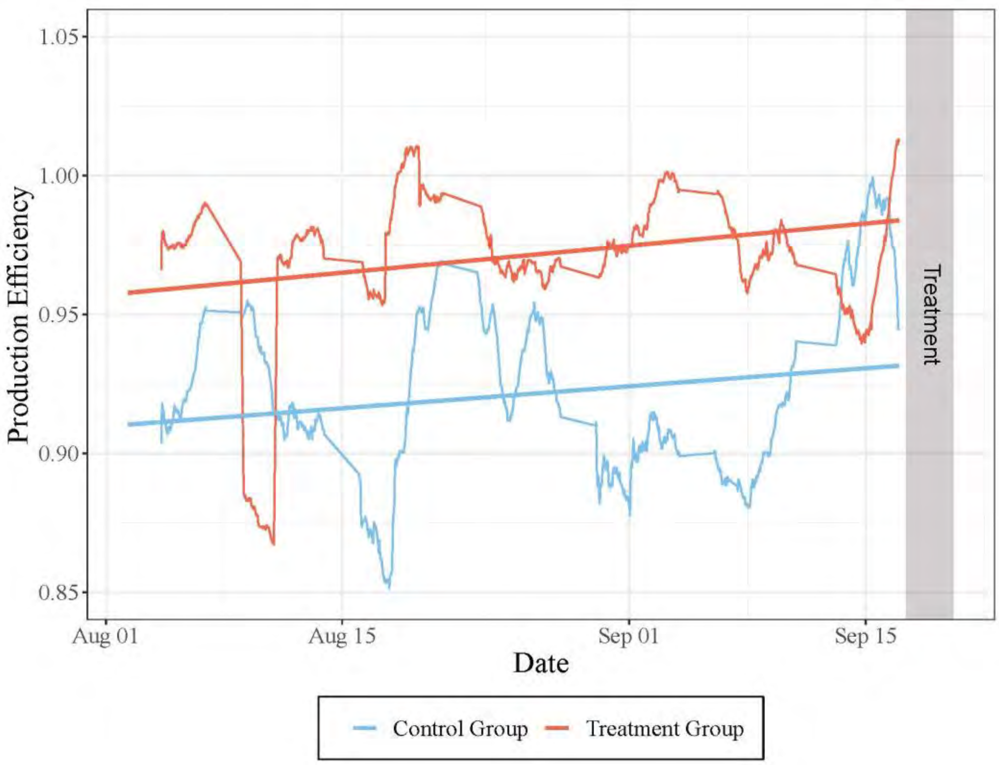
[a]Scores are computed after excluding all unfinished tasks, data stems from the system's back-end.

APPENDIX C: PARALLEL TREND ASSUMPTION FOR MODIFIED TIME PERIOD

C.1 Production Efficiency

Our results (p = .625) suggest that the relative difference of both groups did not significantly change over time before the treatment for production efficiency.

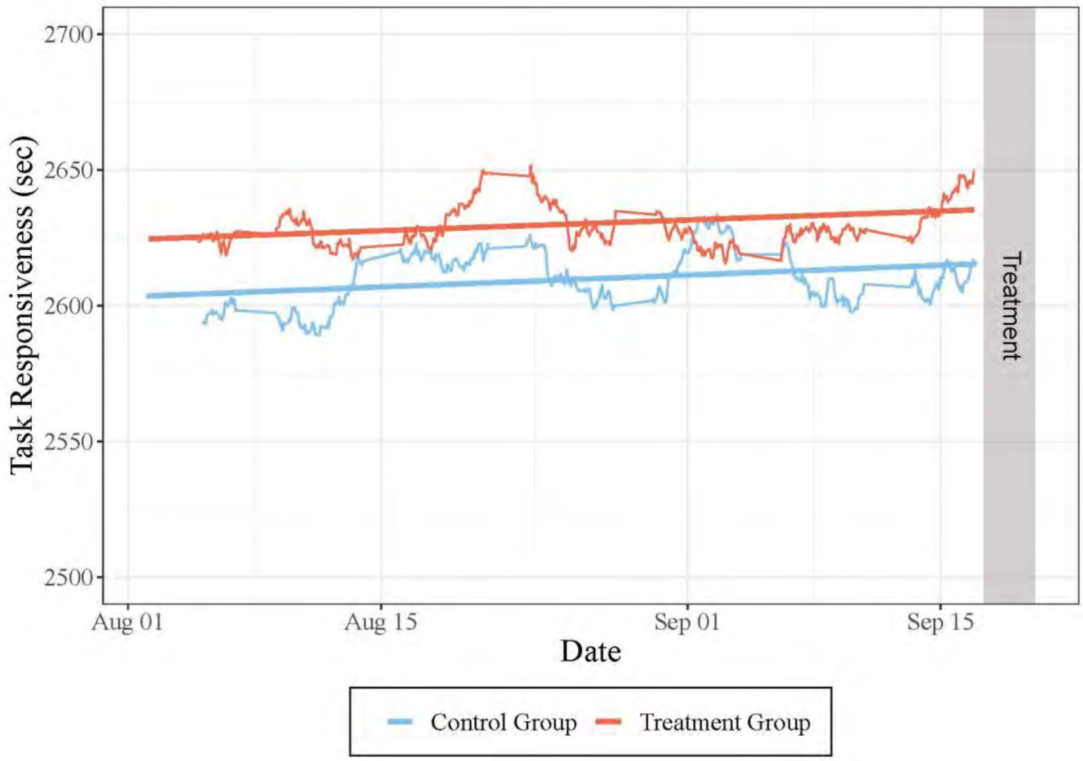
Table 10: Ex-ante Parallel Trend for Production Efficiency



C.2 Task Responsiveness

Our results (p = .689) suggest that the relative difference of both groups did not significantly change over time before the treatment for task responsiveness.

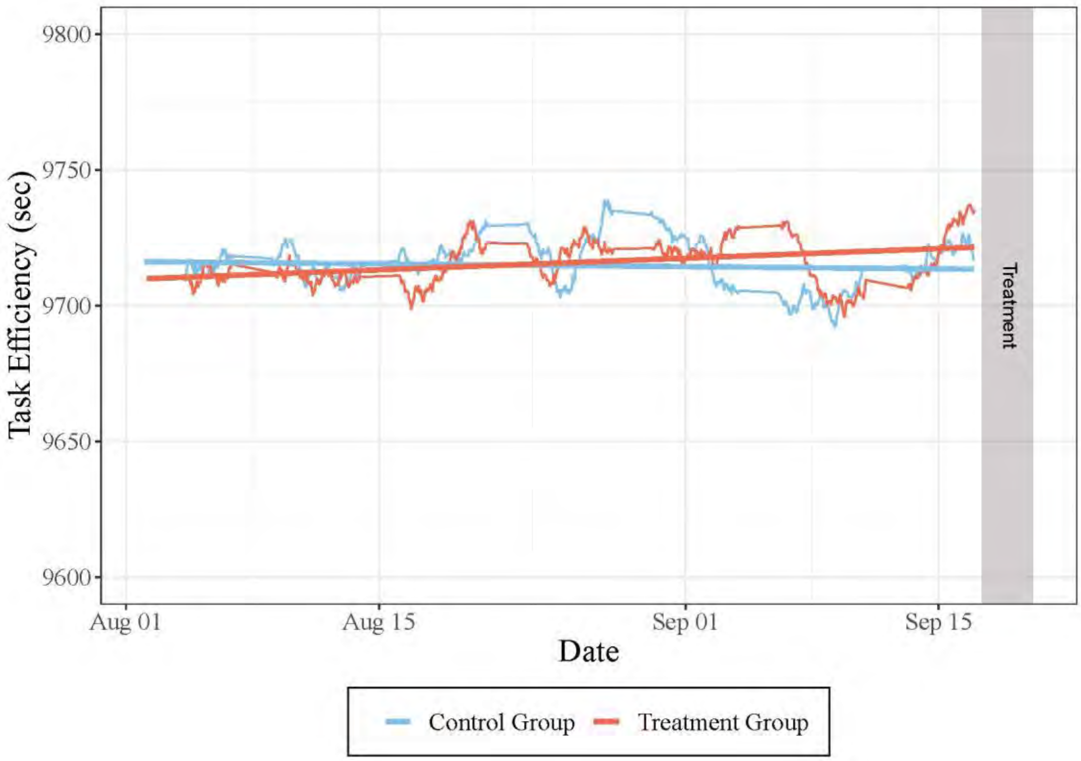
Table 11: Ex-ante Parallel Trend for Task Responsiveness



C.3 Task Efficiency

Our results (p = .218) suggest that the relative difference of both groups did not significantly change over time before the treatment for task efficiency.

Table 12: Ex-ante Parallel Trend for Task Efficiency



APPENDIX D: MACHINE-LEVEL AND TASK-LEVEL ANALYSIS OF MODIFIED TIME PERIOD

Table 13: Machine- and Task-Level Analysis of Impact of choice on Production and Worker Performance

	(1)	(2)	(3)
Dependent Variable:	Production efficiency	Task efficiency	Task responsiveness
Aggregation Level:	Machine-level	Task-level	Task-level
Treated with *choice*	0.031***	12.101	-71.061***
Time fixed effects	Included	Included	Included
Group fixed effects	Included	Included	Included
Constant	0.901***	9688.264***	18292.601***
Adjusted R^2	0.021	0.003	0.019
Observations	46,200	17,032	17,032

Note: Regressions use robust standard errors. ***Significant at the .05% confidence level

APPENDIX E: SHIFT-LEVEL ANALYSIS OF MODIFIED TIME PERIOD

Table 14: Shift-Level Analysis of Impact of choice on Production and Worker Performance

	(1)	(2)	(3)
Dependent Variable:	Production efficiency	Task efficiency	Task responsiveness
Aggregation Level:	Shift-level	Shift-level	Shift-level
Treated with *choice*	0.074***	10.529	-65.792***
Time fixed effects	Included	Included	Included
Group fixed effects	Included	Included	Included
Constant	0.913***	865.349***	506.194***
Adjusted R^2	0.168	0.097	0.136
Observations	286	297	297

Note: Regressions use robust standard errors. ***Significant at the .05% confidence level

DECISION SCIENCES INSTITUTE

Given the high rate of special education teacher attrition, what factors may improve teacher retention?

Cate Brainard
SUNY-Empire State University
catherine_brainard572@esc.edu

ABSTRACT

Retaining special education teachers has been a long-standing challenge and since the outbreak of COVID-19, attrition rates have risen significantly. Research over the past few decades has identified lack of resources, behavioral challenges, conflicts with parents and colleagues, and role-ambiguity as some of the reasons for special education teachers to either leave the field prematurely or retire early. Retaining special education teachers is essential as they not only provide services to vulnerable populations but are also necessary for meeting the mandate of the Individuals with Disabilities Education Act, a federal law.

KEYWORDS: Education, Public Sector, Organizational Relationships

INTRODUCTION

Although much research has been done on the high attrition rates for special education teachers and solutions have been drafted, the problem has gotten worse rather than better. While the research does accurately define the problems faced by special educators, the solutions that are posed are not feasible. The purpose of this study is to answer the following question: Given the high rate of special education teacher attrition, what factors may improve teacher retention? When researching the literature, it supports the importance of mentorship for special educators when they are beginning their career or moving to a new district. Unfortunately, given the teacher shortage, appropriate mentors are few, making the implementation of the solution not possible. Furthermore, role-ambiguity has led to an expansion of roles and responsibilities for the special education teacher whose job is to evaluate students with disabilities, and design and implement data-based interventions. However, over the past twenty years, the expectations of administrators and parents have far exceeded the parameters of the job. While there is ample research about the problem, those most qualified to identify a solution are those who are closest to the problem: special education teachers.

LITERATURE REVIEW

How to hire and retain effective special education teachers has been an issue for several decades and pivoting to online instruction at the onset of the COVID-19 pandemic exacerbated issues that had been long simmering in the field. A press release by the National Center for Education Statistics, stated that “public schools report they are struggling with a variety of staffing issues, including widespread vacancies, and a lack of prospective teachers…[and] of schools reporting at least one vacancy, special education was identified as the teaching position with the most vacancies, with 45 percent of schools reporting this vacancy” (*Press Release - U.S. Schools Report Increased Teacher Vacancies Due to COVID-19 Pandemic, 2022*). Because service hours on the Individualized Education Plan (IEP) are legally binding, special educators must deliver specialized instruction remotely to students with disabilities which proved challenging. Consequently, when studying disparities in service delivery during the pandemic, of parents and guardians "report[ing] that their child had been receiving services or accommodations through an individualized education plan or 504 plan pre pandemic, [only] 61% of those students reported continuing to receive these services in May" (Haderlein et al., 2021). Bonnie Billingsly, an expert on teacher retention, writes that "special educators must manage many responsibilities (e.g., instruction, resources, time management, bureaucratic responsibilities) and work with administrators, general educators, and parents who may not understand what they do or support their efforts" (Billingsley, 2004). This is the manifestation of the role-ambiguity that Billingsly wrote about in an article for the Journal of Learning Disabilities in 2004. Like many other issues of access that were illuminated during the pandemic, this rose to a critical level. Many colleagues in the field worked ten-hour days online with students, parents, administrators, and the content teachers who were posting work, but not providing instruction. The responsibility of instructing the students with learning disabilities fell directly on the special educator who had to deliver instruction across all content areas. Special education became a panacea for all the social, emotional, and academic issues brought on by the pandemic, further blurring the boundaries of the position. Furthermore, online instruction during the spring of 2020 led to learning gaps and increased levels of anxiety that left many students behind when they returned to school in the fall. These learning gaps manifested in what appeared to be an inability to access the curriculum. Referrals to special education spiked, forcing teachers into meetings that often resulted in an eventual placement in special education services. This contributed to an already existing problem as researchers David Peyton and Kelly Acosta consider that "a special education teacher is pulled in multiple directions, with little time to provide instruction… [and] that special education teachers spend as little as 40 percent of their day actually teaching, with paperwork, meetings… [while] roles and responsibilities can change from year to year, limiting their instructional effectiveness" (Peyton & Acosta, 2022). With the increasing caseloads and meetings resulting from the pandemic, it is no wonder that teachers started to leave the field in numbers greater than ever.

Prior to the pandemic, there was already much research addressing special education teacher attrition including the use of "alternative routes to certification that provide little in-the-classroom training" (Podolsky et al., 2019) and while many teachers identify personal reasons or dissatisfaction with assessment protocols, some do return to teaching in different content areas citing the "availability of positions [and] teacher retirement benefits" (Podolsky et al., 2019). Although special educators *were* returning to the field to teach other subject areas prior to the pandemic, there is little to support that this is still the case. Writing for Education Week, Madeline Will addresses the teacher shortage following COVID 19, but points out that, different from other content areas, lowering licensing standards for Special Education teachers is not allowed under the IDEA. While she acknowledges that this contributes further to the shortage, Will cites evidence that "special education teachers who are fully prepared are more likely to stay in the classroom for longer and are more likely to improve their students' academic outcomes" (Will, 2022). In an article for the Council for Exceptional Children, Kevin Monnin, et al. view the high rates of teacher attrition through the lens of policy and how the political climate has made funding for public education more challenging. The problem is further exacerbated by COVID-19 as more students are identified as needing special education services and they wrote that "as a consequence of the COVID-19 pandemic, education has moved to the forefront of the public consciousness, and with this awareness may come increased attention from policymakers seeking to remedy a decades-old problem" (Monnin et al., 2021). As the shortage of special education teachers is at this critical juncture, perhaps "it can be anticipated that state education agencies will need to use this funding from the COVID-19 Relief Package to mitigate the growing attrition of special educators" (Monnin et al., 2021). Because a free and appropriate education (FAPE) for students with disabilities is a legal mandate, one would think that government agencies would “propose and implement short-term and long-term solutions to remedy the teacher shortage” (Monnin et al., 2021).

In 1975, Congress passed the Education for All Handicapped Children Act (Public Law 94–142) (US Department of Education, 2019). This law mandates that all students with special needs be provided with “appropriate special education and related services, and aids and supports in the regular classroom” (US Department of Education, 2019). Furthermore, there is a provision that those who are implementing the services be well trained, receive appropriate professional development and have the “skills and knowledge necessary to improve the academic achievement and functional performance of children with disabilities, including the use of scientifically based instructional practices, to the maximum extent possible” (US Department of Education, 2019). Given that this is the law, what is the solution to the long-standing problem of retaining special educators; one that has been exacerbated by the pandemic? One phrase that reappeared across the literature was teacher induction, the concept of how a teacher is supported as they adapt to a position in a building. Vickie Donne and Fan Yu Lin identify areas that special educators need support including “collaboration, mentoring, and assistance applying skills” (Donne & Lin, 2013). Additionally, developing and implementing strategies for completing the excessive amounts of paperwork is also an area of need. Because Individualized Education Plans (IEPs) are legally binding, special educators are tasked with keeping large amounts of data as well as documentation that they are following the IEP. There are ongoing reminders that their methods of service delivery must be defensible in a hearing if one were to question their practice. One element of induction is the mentoring process as “a preponderance of the research about teacher retention and attrition suggests that lack of or little support for new teachers might contribute to a decision to leave the profession” (Madigan & Schroth-Cavataio, 2012). Despite the stressors, however, there is also evidence that it is the collegial relationships and good mentors that make an impact on one's choice whether to leave the profession or not.

Madigan et al., did their own literature review of qualitative studies of the impact that peers can have on an educator’s decision to stay in the field including one study of “in which 156 teachers were surveyed reported that other special education teachers provided the most help for them in their new jobs” (Madigan & Schroth-Cavataio, 2012). However, when teachers have high caseloads, and excessive paperwork and meetings, mentoring a new teacher is not likely to be an additional task that they are willing to take on. To foster mentorships, workloads would have to be reduced, but that would take a financial commitment. David Peyton and Kelly Acosta express it best when they write that “state and district leaders will first need to identify the scope of the problem. Then they can marshal the data and resources necessary to act. Each [strategy] for… retaining special education teachers requires a substantial investment of time, resources, and coordination among a host of stakeholders” (Peyton & Acosta, 2022). However, as long as administrators do not listen to the teachers, and boards of education do not budget for additional supports, the problem will continue to exist.

THEORETICAL MODEL & HYPOTHESIS

I surveyed ten special education teachers, from two high schools in Connecticut, and one first year middle-school teacher from a school in Los Angeles. A review of the literature has led me to hypothesize that role-ambiguity leads to job related stressors resulting in high-special ed teacher attrition. Utilizing a mixed-methods research design detailed in Research Methods In Public Administration and Nonprofit Management (McNabb, 2017), I drafted a survey that includes two rank choice scales, a rating scale and two narrative responses. The first question asks respondents to rank the stressors that have been categorized either as related to role-ambiguity or typical job stressors. The numerical value will be calculated and the category that scores highest will be coded accordingly. The second question asks respondents to rank the rewards of the position from greatest to least and have been categorized as either a non-mitigating factor or a mitigating factor to the stressors. The numerical value will be calculated and the category that scores higher will be coded accordingly. Finally, the third question asks the respondents to rate how often they think or say statements related to wanting to leave or remain in the job. Again, these statements have been categorized as being negatively or positively correlated with high teacher attrition. For this question, the numerical rank will be calculated and the category that has the higher value will be coded accordingly. (See Appendix A.) Using the quantitative data, I will evaluate using an interaction model that shows the relationship between stressors caused by role ambiguity, perceived rewards of the job, and teacher attrition.

The two short narrative responses pose the questions “In what ways has the COVID-19 pandemic impacted your position?” and “In an ideal world, what changes could be made to improve your job satisfaction?” and are added to provide qualitative data for greater insight into the stressors of the position. This approach has its roots in the ethnographic process in that I am a member of the group being studied (McNabb, 2017, 321) so I have experienced and observed the stressors and rewards identified in the statements included in the ranking and rating scales. I found that while there is a lot of research about teacher attrition, not as much written by those *directly* in the field, which may explain why, as stated earlier, the issue of retention and attrition has only gotten worse. The two short narrative-based questions are purposely left open-ended. The first, “In what ways has the COVID-19 pandemic impacted your position?” was written so as not to limit responses to a series of set ideas, rather to allow for the free expression of each individual’s experience. The second “In an ideal world, what changes could be made to improve job satisfaction?” is also open-ended for the same reasons as the first and is prefaced with “In an ideal world…” because the morale is so low. I did not ask directly what can be done to reduce special education teacher attrition rates because that presents as an overwhelming problem, rather, my process is to gather information about what my colleagues consider to be problems and what can be done for them individually.

For the consent, I utilized a template found on Cornell University’s Research Services webpage (*IRB Consent Form Templates | Cornell Research Services*, 2019) and is linked on the Survey in Appendix A. Potential limitations to this study are the small number of participants and that the study is directed at secondary-level special education teachers who teach primarily students with learning disabilities. Special education certification covers kindergarten through twelfth grade and encompasses students with intellectual, emotional, and learning disabilities. It is a broad-spectrum certificate with different challenges depending on the population, therefore I am limiting my scope to this specific population of teachers.

Utilitarian Theory of Decision Sciences

When viewing the problem of high attrition through the lens of utilitarianism, the costs of not providing enough resources while incrementally adding additional responsibilities benefits no one. The system as it stands is economically unsustainable as it wastes both financial and human capital. Utilizing the theory of decision sciences, an approach that engages all the stakeholders and takes a global view of the problem would enable decision makers to parse out what needs to change and how.

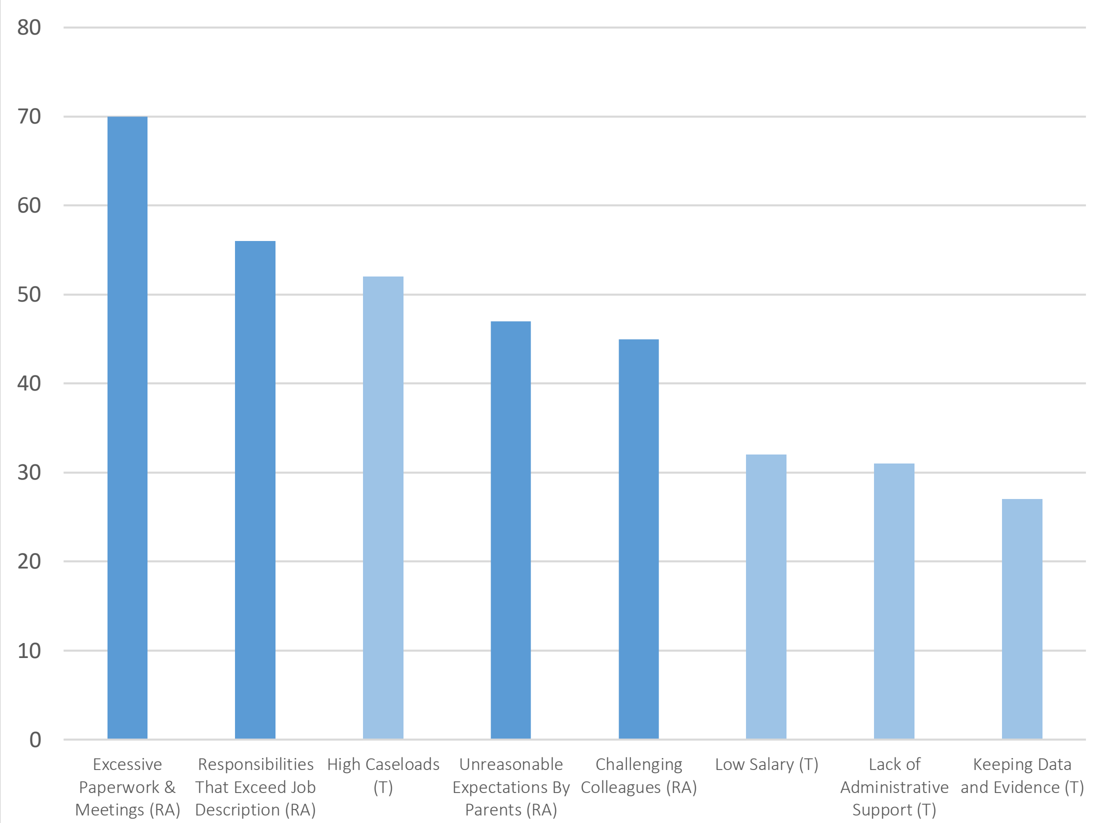
<u>Utilitarian Theory of Decision Sciences: Key Assumptions and Tenets</u>

Using the data gathered, I seek to prove that there is a relationship between stressors caused by role-ambiguity and teacher attrition while mitigating factors may improve teacher retention when the perceived rewards like “positive collegial relationships” and “salary and benefits” are ranked higher. Moreover, I seek to prove that the qualitative data collected from the two narrative questions will provide empirical evidence that COVID-19 has exacerbated job-related stressors, but those who remain in the field are motivated to generate viable solutions.

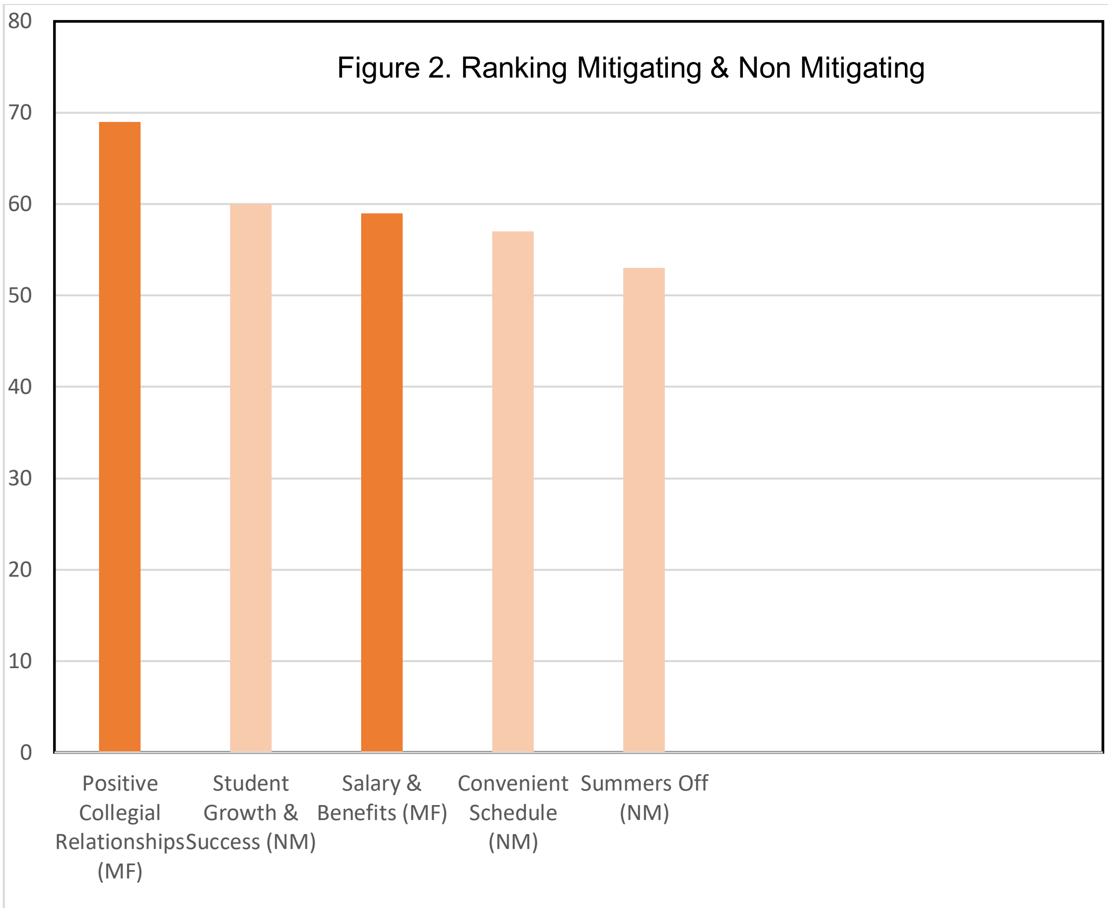
RESULTS OF THE STUDY

Stressors associated with the role of special educator are specified in Figure 1 and are categorized as either associated with role ambiguity (RA) or typical (T) stressors. This is an important distinction because many of us go into the field cognizant of the typical stressors but are not prepared for the expansion of responsibilities that our jobs have evolved to. When determining what variables have a greater impact on the stress of the special educator, the results indicated that it is those related to role ambiguity.

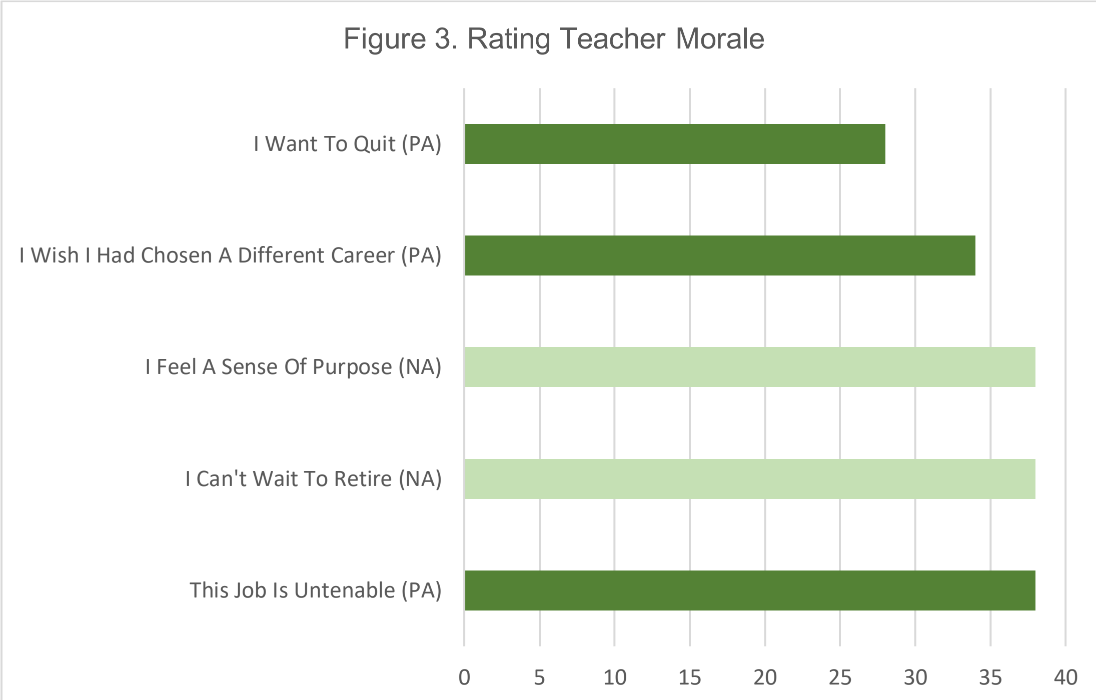
Figure 1. Causes of Special Education Teacher Attrition



In Figure 2, mitigating and non-mitigating variables are ranked to determine how special educators feel about the benefits of the job and if they can outweigh the physical and emotional costs. Positive relationships and salary and benefits are seen as mitigating because as humans, if we are among those with whom we feel connected and who share a common experience, we are more likely to stay. Furthermore, as teacher salaries increase, the financial security can outweigh the stressors. I ranked student growth and progress, convenient schedule, and summers off as non-mitigating factors because those are variables that may be replicated in other content areas. When averaging the sums of the mitigating factors and the non-mitigating factors, mitigating factors were ranked highest overall.



In Figure 3., statements rated by the participants from very frequently (1-2 times per week) to never are represented. They are categorized as either positively correlated (PA) with teacher attrition or negatively correlated (NA). Statements like “I want to quit” and “this job is untenable” are clear examples of being associated with teacher attrition, however, when averaged out, statements negatively associated with attrition like “I feel a sense of purpose” were ranked higher. While the statement “I can’t wait to retire” may appear to be positively associated, people who are closer to retirement will generally stay because they may want to save all they can for their upcoming retirement.



Using the Interaction model, I will determine the impact of mitigating factors and perceived rewards on stressors, to identify possible changes needed to address this near dire problem. Role Ambiguity, the dominant category of stressors has a value of one, whereas mitigating factors have a value of zero as they solve for teacher attrition. When we multiply the two, we get zero meaning that teachers who feel compensated and have solid collegial relationships are less likely to end their career before retirement. Given that the same teachers who rated their salary and relationships also completed the rating scales, they ranked variables negatively associated with attrition higher.

When asked to respond to short answer questions about the impact of teaching remotely during COVID-19, some of the responses included being "expected to be always "on" and accessible and frustration with mainstream teachers "relying more on digital communication which is often unclear and time consuming to wade through." Other issues included increase in special education referrals, which often results in additional meetings and paperwork, as well as an "increase of students with mental health disorders and students regressing." One participant summed it up aptly, writing "many of my students are dealing with gaps in learning and having a hard time managing stressful situations. The mental health of my students has decreased. Students are having a hard time being held accountable and have a limited ability to receive constructive feedback."

When asked what changes could be made to improve job satisfaction, there was a common theme among all participants and included: "fewer responsibilities, especially those which do not include interacting with students…[and] our colleagues held to the same standard that we are." Interestingly, the one teacher from California wrote that "more needs to be taken off our plates instead of things constantly being added to them. It's always a game of catch up and constant anxiety about workload management. Smaller caseloads would also make a world of difference." Another teacher responded "In ideal world, I would change how special education teachers are introduced into the education field. There is a plethora of skills that needs to be learned on the job but dealing with both legal paperwork and trying to balance a schedule with no support from administration can be daunting…I wish there was also more prep time especially when planning for multiple grade instruction. These changes are more district specific but, in its entirety, I would change the way new special educators are trained." This essentially encapsulates all the other responses into one.

CONCLUSION

While I have 25 years of experience in the field, it is with curiosity that I set out to perform this study. Special education is a federal mandate intended to address the needs of a protected class, and it is essential that it be done with fidelity. Reducing stressors associated with role ambiguity has been an ongoing challenge and one that has been further exacerbated by COVID-19. Although much of the literature supports the idea that having a strong mentor is what most teachers are seeking early in their careers, I posit that if the special educator's roles and responsibilities were clear, this would reduce the number of meetings and amount of paperwork that takes us from doing what we set out to do: teach. Although this study is small, the evidence substantiates the significant impact of role-ambiguity on job related stressors, while factors such as positive collegial relationship can mitigate the impact of the stressors. Contractually, special education teachers are responsible to evaluate students with disabilities and then to design and implement scientifically based strategies needed to mitigate the manifestation of the identified disability. However, role-ambiguity has left the special educator responsible for fulfilling many other roles for their students, leaving little time to meet the requirements of IDEA putting both the teacher and the district at risk of non-compliance.

Appendix A

Special Education Teacher Survey (Adapted) Special Education Teacher Survey Consent Form for Study
Considering the daily stressors of a special education teacher, rank the following from most to least stressful (quantitative) 1. Lack of administrative support 2. Responsibilities that exceed job description 3. Unreasonable expectations by parents 4. Challenging colleagues 5. Excessive paperwork & meetings 6. High caseloads 7. Low Salary 8. Keeping data & evidence
Considering the rewards of being a special education teacher, rank the following from greatest to least (quantitative) 1. Positive collegial relationships 2. Student growth & success 3. Salary & benefits 4. Convenient schedule 5. Summers off
Rate the following statements from Very Frequently to Never (quantitative) (very frequently=1-2 a week; frequently=1-2 a month; sometimes=1-2 a year; rarely; never) 1. I wish I had chosen a different career 2. I can't wait to retire 3. I stay because I feel a sense of purpose 4. I feel that this job is untenable 5. I want to quit
Respond to the following: In what ways has the COVID-19 pandemic impacted your position? (qualitative)
Respond to the following: In an ideal world, what changes could be made to improve job satisfaction? (qualitative)

Role-Ambiguity (RA coded 1) / Typical job-related Stressors (T coded 0)

Non-Mitigating factors to stressors (NM coded 1) / Mitigating factors to stressors (MF coded 0)

Positive relationship/ teacher attrition (PA coded 1) / negative relationship/teacher attrition (NA coded 0)

REFERENCES

Billingsley, B.S. (2004). *Promoting Teacher Quality and Retention in Special Education,*. Journal of Learning Disabilities. https://doi-org.library.esc.edu/ 10.1177/00222194040370050101

Donne, V., & Lin, F.-Y. (2013). Special Education Teacher Induction: The Wiki Way. *Taylor & Francis Group*, *86*(2), 43-47. Retrieved April 17, 2023, from: https://www.jstor.org/stable/43999315

Haderlein, S., Saavedra, A. R., Silver, D., Polikoff, S., Rappaport, A., & Garland, M. 2021, August 23). Disparities in Educational Access in the Time of COVID: Evidence From a Nationally Representative Panel of American Families. https://doi.org/10.1177/23328584211041350

IRB Consent Form Templates | Cornell Research Services. (2019). Cornell Research Services. Retrieved April 21, 2023, from https://researchservices. cornell.edu/forms/irb-consent-form-templates

Madigan, J. B., & Schroth-Cavataio, G. (2012). *Mentorship of Special Educators*. SAGE Publications.

McNabb, D. E. (2017). *Research Methods for Public Administration and Nonprofit Management*. Routledge.

Monnin, K., Day, J., Strimel, M., & Dye, K. (2021, June 1). *Why Now Is the Perfect Time to Solve the Special Education Teacher Shortage*. Council for Exceptional Children. Retrieved April 15, 2023, from https://exceptionalchildren.org/blog/why -now-perfect-time-solve-special-education-teacher-shortage

Peyton, D., & Acosta, K. (2022, January 20).Understanding Special Education Teacher Shortages. DigitalOcean. Retrieved April 17, 2023, from https://nasbe. nyc3.digitaloceanspaces.com/2022/01/Peyton-Acosta_Jan-2022-Standard.pdf

Podolsky, A., Kini, T., Darling-Hammond, L., & Bishop, J. (2019, April 8). *Strategies For Attracting and Retaining Educators: What Does The Evidence Say?* Education Policy Analysis Archives. https://epaa.asu.edu/index.php/epaa/article/view/3722/2231

Press Release - U.S. Schools Report Increased Teacher Vacancies Due to COVID-19 Pandemic, New NCES Data Show - March 3, 2022. (2022, March 3). NCES. Retrieved April 18, 2023, from https://nces.ed.gov/whatsnew/press_releases/3_3_2022.asp

US Department of Education. (2019). Individuals with Disabilities Education Act. Retrieved April 19, 2023, from https://sites.ed.gov/idea/statute-chapter-33/subchapter-i/1400

Will, M. (2022, October 25). *States Are Desperate for Special Ed. Teachers. But They Can't Cut Corners to Get Them*. Education Week. Retrieved April 17, 2023, from https://www.edweek.org/teaching-learning/states-are-desperate-for-special-ed-teachers-

DECISION SCIENCES INSTITUTE
Heart Disease Prediction Using Feature Selection and Classification Models

Dinesh K. Sharma
University of Maryland Eastern Shore
dksharma@umes.edu

Tarun Dhar Diwan
Atal Bihari Vajpayee University
tarunctech@gmail.com

Ajay Tiwari
Atal Bihari Vajpayee University
officialajay2@gmail.com

ABSTRACT

Heart disease is the leading cause of mortality worldwide, and hospitals provide a plethora of medical information about it. In this research work, we conducted an operational evaluation of models constructed using ten feature selection approaches to improve heart disease accuracy and found that the Backward Feature Selection (BFS) procedure had the highest levels of classification accuracy (89%), precision (91%), sensitivity (81%), and f-measure (86%), compared to the previous studies. This study suggests that BFS is a promising feature selection technique for improving the accuracy of classification models for heart disease. Future studies should investigate its use with other classification methods and data sets.

KEYWORDS: Heart disease, Feature selection, Classification, Classification accuracy, Backward Feature Selection (BFS)

INTRODUCTION

The information age has led to the proliferation of computer-aided systems, generating vast amounts of data and strengthening new power structures. Extracting valuable information from this data is a daunting challenge for researchers and practitioners. Machine Learning (ML) techniques, particularly in the Data Mining field, are emerging and promising in identifying connections and determining significant facts through advanced statistics. Data mining techniques can help doctors make more accurate diagnoses of patients' conditions, such as heart diseases, malignancies, diabetes, skin disorders, and renal diseases. Cardiovascular diseases have a high death rate worldwide, with the World Health Organization estimating that the number of deaths due to cardiovascular diseases will reach 30 million by 2040 (Patro et al., 2022). To reduce the costs associated with diagnosis and treatment, data mining techniques can help determine a patient's risk of developing heart issues during an early period.

High-dimensional and heterogeneous data analysis is necessary for analyzing various illnesses. Model overfitting and dimensionality issues can be addressed by decreasing dimensionality and selecting features efficiently. The research aims to measure the usefulness of various feature selection algorithms. The real datasets were obtained from Cleveland datasets accessible at the University of California. Experimental techniques were used to conduct thorough tests on real datasets related to heart disease and evaluate their performance metrics, including accuracy, precision, and recall.

LITERATURE REVIEW

The reviewed studies employ various methods, such as tree-based ensemble classification techniques, neural networks (NN), Convolutional NN (CNN), K-Nearest Neighbor (KNN), support vector machines (SVM), and Internet of Things (IoT) based patient monitoring. Additionally, various feature selection techniques are utilized for selecting relevant features.

Table 1: Review of literature

Author (year)	Type of Research	Method Used
Aldossary et al. (2022)	Comparative Study	Tree-Based Ensemble Classification Techniques
Almansour et al., (2019)	Comparative Study	SVM and Neural Network
Angayarkanni et al. (2020)	Experimental Study	Feature Selection Techniques
Baviskar et al. (2023)	Experimental Study	Hybrid Deep Learning Classification Models
Cenitta et al. (2021)	Experimental Study	ML Algorithms
Dissanayake and Johar (2021)	Comparative Study	Feature Selection Methods
Dutta et al. (2020)	Experimental Study	Convolutional Neural Network
Elavarasan et al. (2020)	Experimental Study	Hybrid Feature Extraction Techniques
Ghosh et al. (2021)	Experimental Study	Feature Selection Methods with ML Algorithms
Hasan and Bao (2020)	Comparative Study	Feature Selection Algorithms
Latha and Jeeva (2019)	Experimental Study	Ensemble Classification Techniques
Mohanaprakash et al. (2023)	Experimental Study	Artificial Intelligence for Early Detection
Moorthy and Gandhi (2020)	Experimental Research	ANOVA-based whale optimization
Moturi et al. (2021)	Experimental Research	Machine Learning Techniques
Panda et al. (2019)	Experimental Research	Feature Selection, Prediction Systems
Patro et al. (2022)	Experimental Research	Correlation, Feature Selection Techniques
Pemmaraju et al. (2022)	Experimental Research	Feature Selection, Machine Learning Techniques
Phiri et al. (2022)	Experimental Research	Machine Learning Techniques
PraneethKumar et al., (2021)	Comparative Study	Machine Learning Techniques
Rosely et al. (2019)	Review Paper	Fish Swarm Algorithm
Das et al. (2022)	Experimental Research	Machine Learning Methods
Sabab et al. (2016)	Experimental Research	Classification, Feature Selection Technique
Senan et al. (2021)	Experimental Research	Classification, Feature Selection Technique
Shree et al. (2022)	Experimental Research	Machine Learning Techniques
Solorio-Fernández et al. (2020)	Review Paper	Unsupervised Feature Selection Technique
Spencer et al. (2020)	Experimental Research	Classification, Feature Selection Technique
Takci (2018)	Experimental Research	Feature Selection Methods
Tasnim and Habiba (2021)	Experimental Research	Data Mining Techniques, Feature Selection
Usha and Kanchana (2022)	Experimental Research	Data-Driven Approach, Feature Selection Techniques

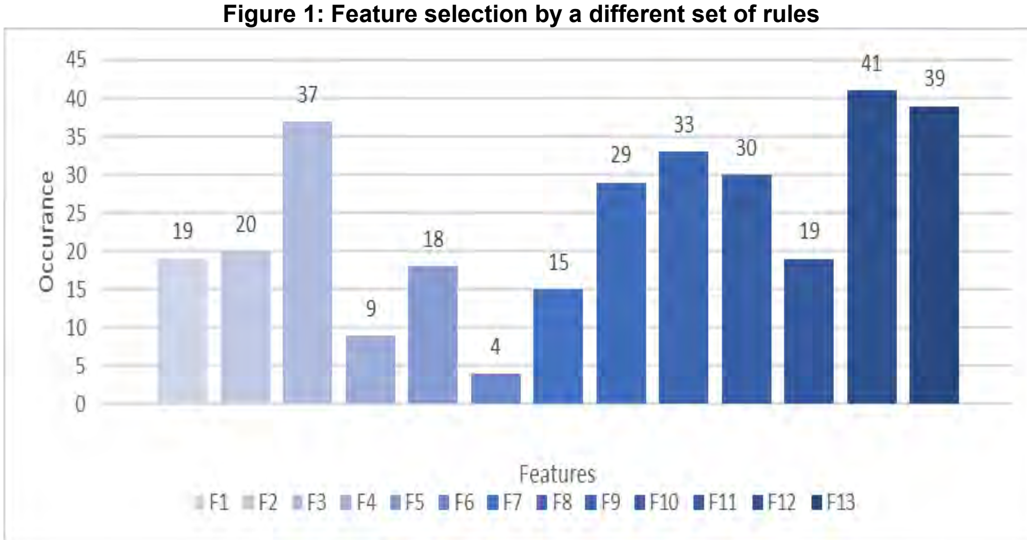
FRAMEWORK OF USING FEATURE SELECTION TECHNIQUES

The Proposed method aims to improve classification perfection by reducing variables in a dataset on cardiovascular sickness, using diagnostic criteria and data splitting, and dividing into segments. Preprocessing the dataset involves removing absent values, applying StandardScaler and MinMaxScaler, and applying missing value treatment. The feature selection process is used to accurately detect heart illnesses and create a more accurate model by excluding or underrepresenting less significant elements. Different feature selection strategies, classified as filter, wrapper, or embedding, are applied to the original datasets, with the initial subset being the first step in selecting the appropriate approach.

Feature Selection by Various Filtering Techniques

Filter-based selection processes use statistical approaches to determine the connection between self-reliant and reliant features, input and target attributes. This method evaluates features based on their resemblance to data appearances and does not rely on learning methods. Features are selected based on output results from various statistical procedures. Figure 1 shows a rundown of the appearances that were prioritized by the algorithm used to choose features. Its primary function is to quantify the features' appearances using various measuring criteria. In order to choose the proper subset, this collection of qualities is evaluated in terms of their degree of dependency on the output attribute. Chi-square, MI, ReliefF, and the ANOVA technique are the four filter-based feature selection approaches primarily used in this suggested work.

Figure 1: Feature selection by a different set of rules



Chi-Square: A feature selection method that adheres to the univariate statistical methodology (PraneethKumar et al., 2021). This method aims to determine which features in a dataset are dependent on, or correlate with, the variable of interest. At this point, we should eliminate the appearances unrelated to the target variable and any dependent features that are not deemed very relevant for predicting the target variable.

$$X^2 = \sum \frac{f_O - f_E}{f_E} \qquad (1)$$

If two appearances are considered self-reliant, a lower Chi-square score is produced because the frequency seen is relatively near to the anticipated frequency.

Table 2: A filter-based ranking to determine the order of the appearances

Filtering Techniques	Ranking												
	1	2	3	4	5	6	7	8	9	10	11	12	13
ANOVA	F13	F12	F10	F9	F8	F3	F11	F2	F1	F7	F5	F4	F6
Chi-square	F8	F12	F10	F13	F5	F9	F1	F4	F3	F7	F2	F11	F6
Mutual Information	F13	F12	F9	F3	F10	F8	F11	F5	F6	F2	F1	F7	F4
Relief	F13	F12	F9	F10	F3	F10	F8	F2	F1	F7	F5	F4	F6

Mutual Information (MI): A metric that analyses the interdependence of two random variables is called the mutual information between them. This metric also gauges the degree to which uncertainty is reduced. This may also be stated as "measures the reduction in uncertainty." The actions that need to be carried out to complete this process successfully are outlined in the steps shown below. The following instructions and equations are used throughout computing MI scores. Entropy and conditional entropy are two concepts that may be used to evaluate the usefulness of a specific characteristic in the context of predictive modeling of an unknown class of samples (Phiri et al., 2022). It is possible to express the entropy of $H(X)$ in terms of the values "$x_1, x_2, x_3 \ldots\ldots x_n$" using the formula where "means probability mass function of." The formula for calculating the restricted entropy of the two discrete random variables X and Y is:

$$H\,(X) \;=\; -\sum_{i=1}^{n} P(x_i)\; log_2\, P(x_i) \qquad (2)$$

$$H(X|Y) \;=\; -\sum_{i=1}^{n}\sum_{j=1}^{n} P\left(x_i\,, y_j\right)\; log_2\, \frac{P(x_i)}{P\left(x_i\,, y_j\right)} \qquad (3)$$

$$MI\,(X : Y) \;=\; H\,(X) \;-\; H(X|Y) \qquad (4)$$

where $P\left(x_i\,, y_j\right)$ represents the joint probability where $X \;=\; x_i$ and $Y \;=\; y_j$. Estimates for X and Y mutual information $MI\,(X : Y)$ can be formulated in terms of either X's entropy, H (X), and conditional entropy, $H\,(C|X)$.

ReliefF Algorithm: It is a strategy for learning oversaw features used for binary and multiclass problem areas to highlight the importance of the target variable by weighing each feature and presenting its relevance. This approach is referred to as oversaw feature learning. This method may be completed by following the steps that are outlined in the following paragraphs. The instance compared the values of the appearances pertaining to the closest neighbors and adjusted the weights of each feature. The following equation describes how to calculate the weight vector associated with a feature in binary classification:

$$W_i \,=\, W_i \,-\, (x_i - nearHit_i)^2 \,+\, (x_i - nearMiss_i)^2 \quad (5)$$

where $\mathrm{nearHit}$ refers to the closest occurrence of the same class and $\mathrm{nearMiss}$ refers to adjacent occurrence of a distinct class. Instead of the Euclidean distance (L2), the Manhattan distance (L1) is employed in the ReliefF technique to identify the $\mathrm{nearHit}$ and $\mathrm{nearMiss}$ cases. In addition, the $ReliefF$ approach updates the weight vectors by focusing on the absolute differences between $\mathrm{x_i}$ and $\mathrm{nearHit}$ rather than the square of those differences. This is done in place of the traditional way of calculating the square of such disparities.

ANOVA Method: The Analysis of Variance, more often known as ANOVA, A test of statistical significance that estimates the ratio of modifications, such as the modification that results from comparing two dissimilar models, is known as the variance ratio test (VRT). Additionally, the analysis of variance (ANOVA) is helpful for classification since it generates data with numerical inputs and categorical desired variables. (Sabab et al., 2016). The following is a list of the actions that must be completed in order to finish this procedure. Apply the procedure that is provided below in order to calculate the outcomes of the one-way ANOVA F-test.

$$Changes\ between\ Collections \;=\; \frac{\sum_{i=1}^{n} n_i\,(\bar{Y_\iota} - \bar{Y})^2}{K-1} \qquad (6)$$

$$Changes\ within\ Collections \;=\; \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(\overline{Y_{\iota J}} - \bar{Y_\iota}\right)^2}{N-K} \qquad (7)$$

$$F = \frac{Changes\ between\ Collections}{Changes\ within\ Collections} \qquad (8)$$

In which $\overline{Y_i}$ is the average of the samples from $the\ i^{th}$ collection, n_i is the number of values in the i^{th} collection, $\overline{Y}$ is the total average of data, and the standard deviation of the ith collection's findings, where K is the sum of collections, j is the j^{th} observation in the i group, and N is the sum of findings in the overall sample.

RESULTS AND DISCUSSIONS

The comparative analysis used Python for Jupyter Notebook and a sci-kit feature from Arizona State University, which includes over 40 feature selection algorithms. The Cleveland heart disease dataset contains 303 samples, with 164 showing no heart disease and 139 indicating it. Ten feature selection algorithms were chosen for classification, classified as filters, wrappers, or embedded methods. Six classifiers were used for performance analysis, including Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), K-nearest neighbor (KNN), Lasso Regression (LR), and Gaussian Naive Bayes (GNB). The findings show the accuracy of different feature sets selected using Classification techniques.

.

Figure 2: Findings of Initial Feature Selection (IFS) (with Feature 13)

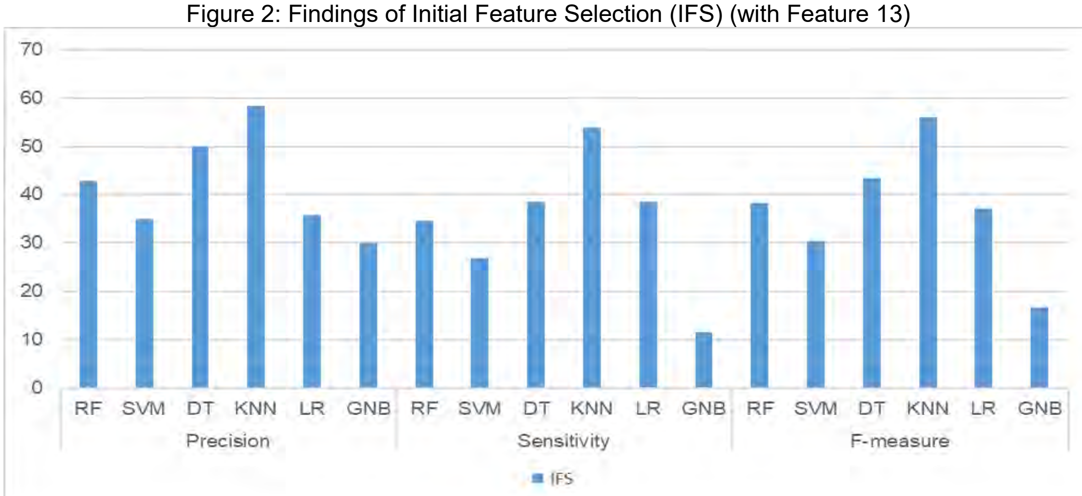


Figure 3: Findings of Chi-square Classification (with Features 6, 7, 8, 9)

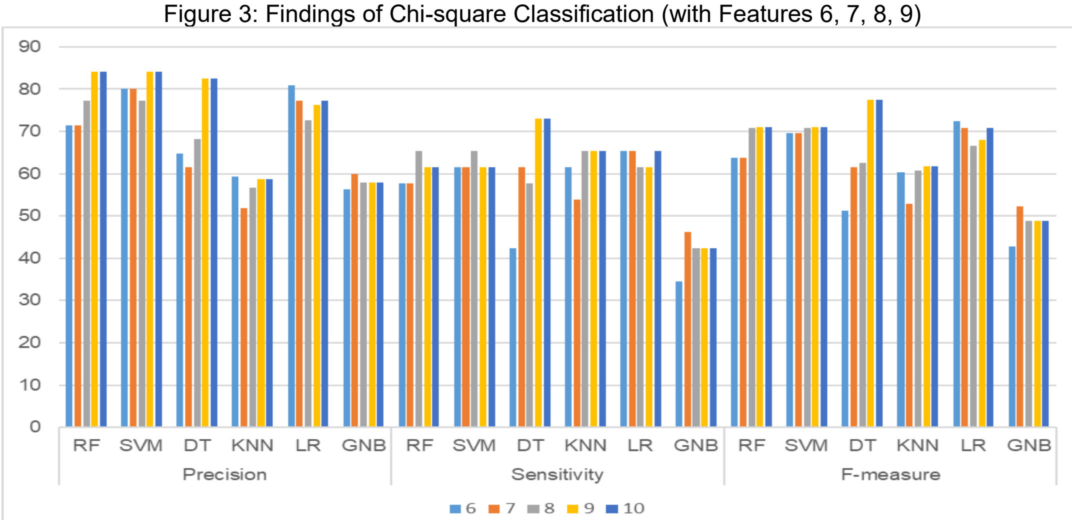


Figure 4: Findings of ANOVA Classification (with Features 6, 7, 8, 9, 10)

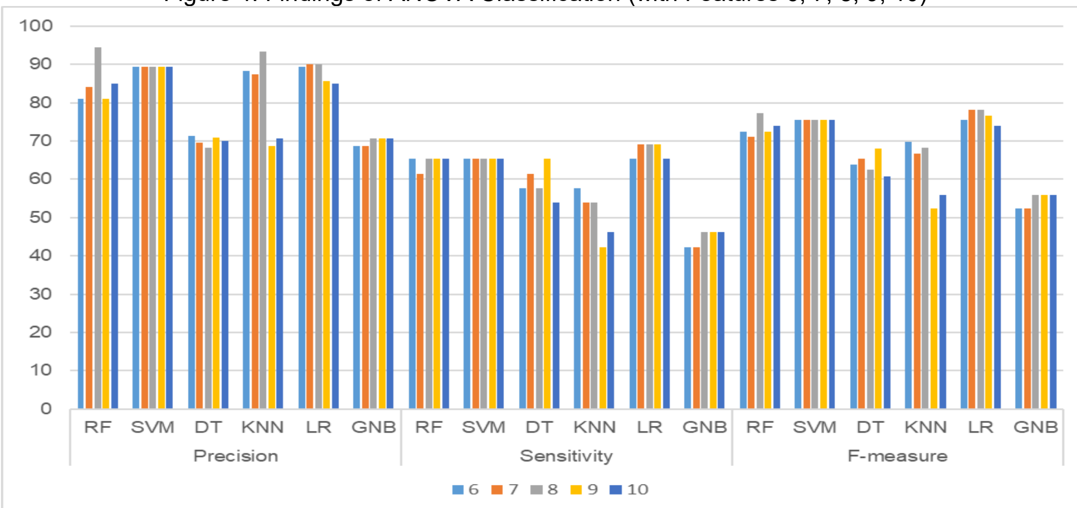


Figure 5: Findings of Mutual Information (with Features 6, 7, 8, 9, 10)

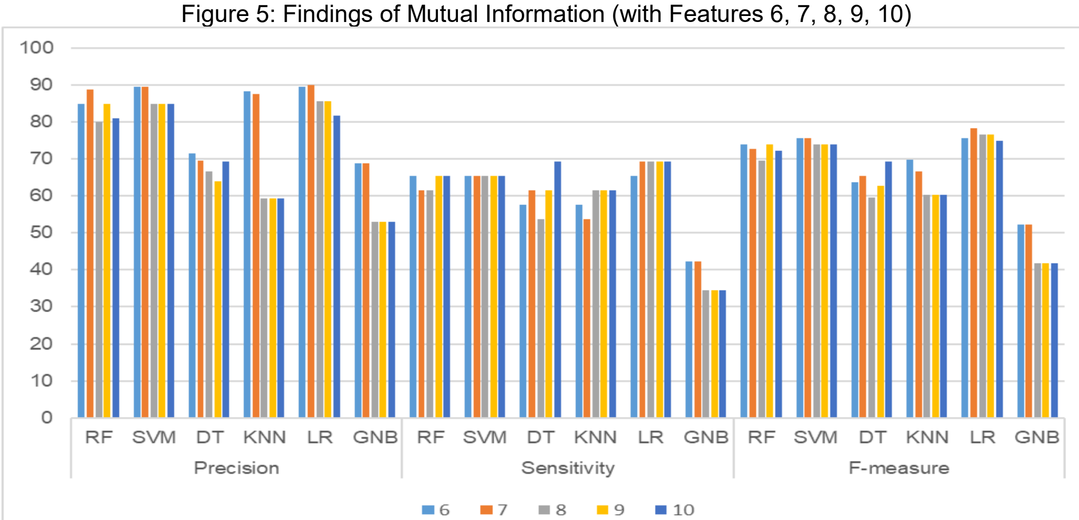


Figure 6: Findings of ReliefF method (with Features 6, 7, 8, 9, 10)

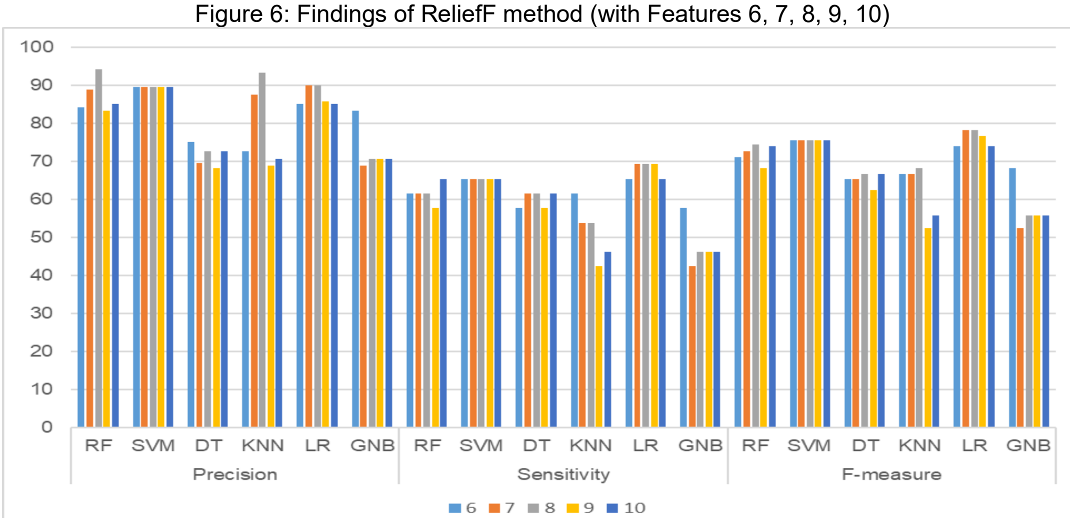
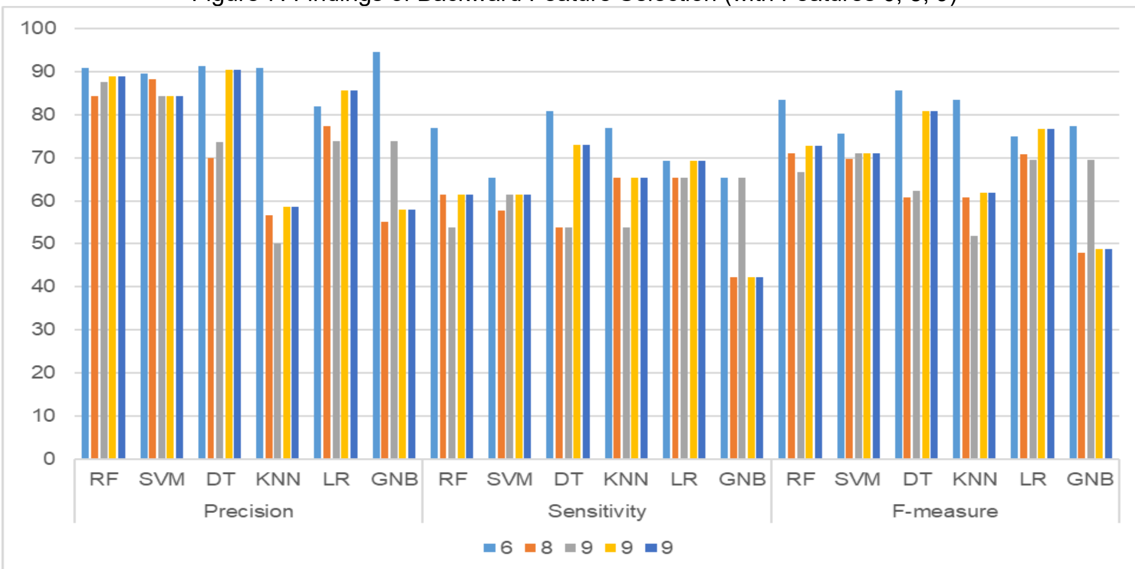


Figure 7: Findings of Backward Feature Selection (with Features 6, 8, 9)



Figures 2–7 show a comparison of many feature selection strategies for predicting cardiovascular disease. RF, SVM, DT, KNN, and LR, as well as GNB, are some of the approaches that have been investigated. In addition, the figures show the results of utilizing each feature selection approach in conjunction with their respective classification algorithms in terms of performance metrics.

CONCLUSION

This study examines the impact of different feature selection methods on heart disease prediction precision. It employs various techniques, including Chi-square, ANOVA, MI, Relief, FFS, BFS, EFS, RFE, LR, and RR, and evaluates six different classification methods. The decision tree performed the best, with the KNN classifier yielding the highest model accuracy of 64%. The study found that feature selection algorithms can effectively detect heart disease using a limited number of appearances. However, determining the optimal approach requires multiple trials and studies. Hybrid techniques using multiple feature-selection algorithms could help extract optimal feature subsets for model development. Utilizing real-time health information from various nations could improve forecasting accuracy and overall performance.

REFERENCES

Aldossary, Y., Ebrahim, M., & Hewahi, N.M. (2022). A Comparative Study Of Heart Disease Prediction Using Tree-Based Ensemble Classification Techniques. 2022 International Conference on Data Analytics for Business and Industry (ICDABI), 353-357.

Angayarkanni, G. (2020). Selection OF features associated with coronary artery diseases (cad) using feature selection techniques. Journal of Xi'an University of Architecture and Technology, 686–689.

Dutta, A., Batabyal, T., Basu, M., & Acton, S. T. (2020). An efficient convolutional neural network for coronary heart disease prediction. Expert Systems with Applications, 159, article ID 113408. https://doi.org/10.1016/j.eswa.2020.113408.

(A complete list of references is available upon request.)

DECISION SCIENCES INSTITUTE

Leveraging Foreign Diversification to Build Firm Resilience: A Conditional Process Perspective

Dominic Essuman
Sheffield University Management School, The University of Sheffield, UK
Email: d.essuman@sheffield.ac.uk
Address: C086, Conduit Road, Management School, Sheffield | S10 1FL, Sheffield, UK

ABSTRACT

This research combines the organizational information processing theory and international business literature to examine how and when foreign diversification relates to firm resilience. The study suggests that while foreign diversification may contribute to firm resilience, foreign market scanning mediates this effect under varying supply chain disruption conditions. An analysis of primary data from 272 SME exporters in Ghana reveals that foreign diversification alone does not explain firm resilience. Instead, the results support the arguments that foreign market scanning positively mediates the foreign diversification-firm resilience relationship, and that this indirect relationship is stronger in highly disruptive supply chain environments.

<u>KEYWORDS</u>: Global Supply Chain Disruptions, Internationalization, Organizational Resilience, Organizational Information Processing Perspective, and international SMEs

INTRODUCTION

The increasing frequency and costs of supply chain disruptions in foreign markets have highlighted the importance of firm resilience for international firms (Pitelis et al., 2023; Lee et al., 2022). These disruptions can lead to reputational damage, inefficiencies, and decreased revenue, making resilience critical for businss survival (Essuman et al., 2023; Manhart et al., 2020). Resilient international firms can endure, adapt, or transform in the face of such challenges (Wieland & Durach, 2021), which can improve their competitiveness and profitability (Iftikhar et al., 2021; Wong et al., 2020).

Although research has advanced the understanding of factors affecting firm resilience, the impact of unique characteristics of international firms on their resilience remains underexplored (see, e.g., Li et al., 2023; Iftikhar et al., 2021; Manhart et al., 2020). Recent attention in international business (IB) literature has turned toward the role of foreign diversification in building resilience (Puhr & Müllner, 2022; Kersan-Škabić, 2021; Kano & Hoon Oh, 2020). Notable events like trade wars and the COVID-19 pandemic have revealed the vulnerability of firms with concentrated foreign market operations during disruptions (Duffy, 2023; Simchi-Levi & Haren, 2022). Accordingly, foreign diversification, operating in multiple foreign markets, has been associated with greater resilience (Puhr &Müllner, 2022). However, this relationship is context-dependent, suggesting the need for further empirical exploration (Puhr & Müllner, 2022).

The ongoing debate on the value of foreign diversification necessitates alternative approaches to analyzing its relationship with firm resilience (Huang et al., 2023). Some view foreign diversification as spreading risks and accessing broad international knowledge, while others see it as a source of complexity and uncertainty (Puhr & Müllner, 2022; Fariborzi et al., 2022; Mondal et al., 2022). While past studies have used contingency and curvilinear analyses in an attempt to resolve conflicting findings, knowledge of the mechanisms and associated boundary conditions linking foreign diversification to specific organizational outcomes is limited (Huang et al., 2023). Accordingly, this research addresses the question: *How and when does foreign diversification contribute to firm resilience?*

This study builds on and extends Puhr and Müllner's (2022) finding that foreign diversification enhances abnormal stock market returns of multinational enterprises during shocks. Focusing of international SMEs, this study argues that internationally diversified firms can benefit from experiential knowledge, financial resources, and adaptive behaviors, which may drive their resiliency (Fariborzi et al., 2022; Puhr & Müllner, 2022). The study further draws on the organizational information processing (OIP) theory to develop a conditional process model to unpack the foreign diversification-firm resilience relationship. The study proposes that while foreign diversification may be associated with complexity and uncertainty (Fariborzi et al., 2022; Puhr & Müllner, 2022), it can stimulate foreign market scanning, which may enable internationally diversified firms to be resilient (Tushman & Nadler, 1978; Srinivasan & Swink, 2018). Foreign market scanning can improve visibility, minimize disruption forecasting errors, and enhance quicker disruption detection and responses (Essuman et al., 2022).

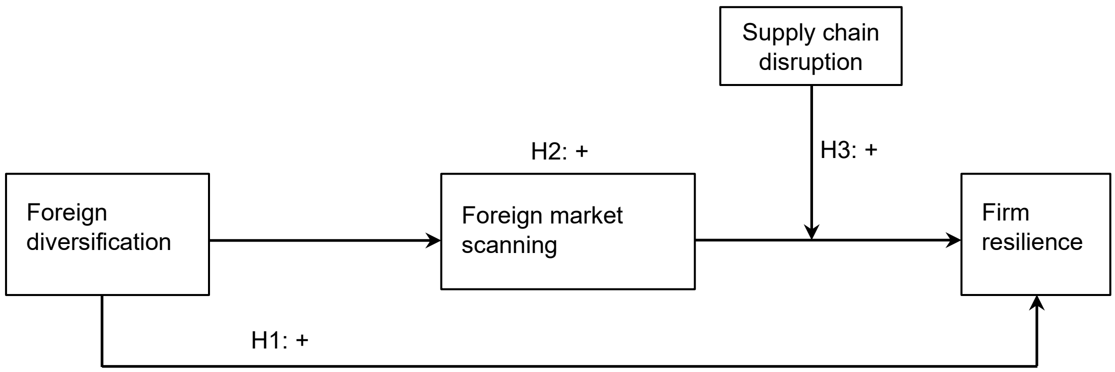
Yet, foreign market scanning is not universally beneficial (Essuman et al., 2022; Bouquet et al., 2009). OIP theory suggests that foreign market scanning may be more effective in uncertain environments (Srinivasan & Swink, 2018; Tushman & Nadler, 1978). Therefore, this study proposes supply chain disruption as a context where foreign market scanning may enhance resilience (Yang et al., 2021; Bode et al., 2011).

The study offers three contributions. First, it advances the limited understanding of the determinants of firm resilience in the context of foreign market operations by explaining the variance in resilience through foreign diversification, foreign market scanning, and supply chain disruption. Second, the study extends Puhr and Müllner's (2022) work by using international SMEs and proposing and demonstrating an alternative approach to unraveling the nuances surrounding the role of foreign diversification in resilience-building. Specifically, the study presents a conditional-process perspective, identifying foreign market scanning as a mechanism through which internationally diversified SMEs, especially those facing greater disruptions, achieve superior resilience. Third, this conditional conditional-process approach helps clarify how and when foreign diversification is more or less beneficial (Huang et al., 2023).

THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

Though there is extensive literature on the consequences of foreign diversification (Arte & Larimo, 2022; Puhr & Müllner, 2022; Fariborzi et al., 2022; Mondal et al., 2022), only recently have scholars begun examining its resilience effects (Puhr & Müllner, 2022). In advancing this line of inquiry, the study first draws on international business literature (Fariborzi et al., 2022; Puhr & Müllner, 2022) to examine the direct relationship between foreign diversification and firm resilience. Next, the study applies OIP theory to propose that foreign market scanning mediates this relationship at varying levels of supply chain disruption (Figure 1).

Figure 1. Conceptual model



Foreign Diversification and Firm Resilience

Foreign diversification, also termed internationalization scope, geographic scope, or multinationality, refers to the extent to which a firm operates in multiple foreign markets (Arte & Larimo, 2022; Puhr & Müllner, 2022; Fariborzi et al., 2022; Mondal et al., 2022). In contrast, firm resilience refers to the capability of the firm to absorb, quickly recover from, adapt to, or transform during supply chain disruptions (Wieland & Durach, 2021; Essuman et al., 2020).

Foreign diversification, though it may have a dark side (e.g., it can induce complexity and increase coordination costs) (Puhr & Müllner, 2022; Fariborzi et al., 2022), tends to be associated with positive market and financial performance outcomes (Arte & Larimo, 2022; Schwens et al., 2018). Notably, Puhr and Müllner (2022) find that diversified multinational corporations exhibit higher resilience. Three key reasons support this evidence.

Firstly, foreign diversification creates new revenue opportunities and stabilizes revenue streams by allowing firms to share operational and market risks across multiple geographical locations (Schwens et al., 2018). Economies of scale associated foreign diversification may further reduce operational costs (Puhr & Müllner, 2022; Fariborzi et al., 2022). Such financial advantages that may result from foreign diversification can enable firms to build buffers against supply chain disruptions through alternative suppliers, diversified logistics, and excess inventory (Puhr & Müllner, 2022; Essuman et al., 2022).

Secondly, internationally diversified firms accumulate experiential knowledge across multiple markets (Arte & Larimo, 2022; Asmussen et al., 2022), reinforcing learning and resilience (Essuman et al., 2023; Bode et al., 2011). As international firms transfer knowledge between markets, their ability to address market-specific disruptions grows (Fariborzi et al., 2022; Arte & Larimo, 2022), which may benefit their responses to disruptions (Bode et al., 2011).

Lastly, operating in numerous markets necessitates adaptive behaviors, fostering agility in responding to disruptions (Huang et al., 2023; Puhr & Müllner, 2022; Mondal et al., 2022). Firms with such behaviors can explore and experiment with novel disruption management solutions (Essuman et al., 2023).

In conclusion, foreign diversification empowers firms with revenue opportunities, knowledge, and adaptive behaviors, which may benefit firm resilience.

H1: Foreign diversification has a positive relationship with firm resilience.

Mediating Role of Foreign Market Scanning

Foreign diversification presents complexity and uncertainty, which can undermine its resilience benefits (Puhr & Müllner, 2022; Mondal et al., 2022). Operating in diverse environments exposes firms to various risks and challenges (Mondal et al., 2022; Fariborzi et al., 2022; Lee & Chung, 2022). Accordingly, we propose that foreign market scanning acts as a key mechanism for internationally diversified firms to navigate such environmental issues and achieve resilience (Yang et al., 2021; Gu et al., 2021).

OIP theory views firms as information processing systems that manage uncertainty by increasing information search and processing activities (Galbraith, 1974; Tushman & Nadler, 1978). Environmental complexities and fluctuations drive uncertainty, which may undermine organizational stability (Bode et al., 2011; Manhart et al., 2020). Therefore, internationally diversified firms are inclined to intensify information search activities to enable them operate deterministically while ensuring operational continuity (Tushman and Nadler, 1978). Foreign market scanning, involving gathering information about overseas market issues (Yu et al., 2019), can aid firms to better understand disruptions and profile their risk levels. These benefits of foreign market scanning can help firms develop and deploy appropriate disruption-response measures (Essuman et al., 2022; Yu et al., 2019).

Increased foreign market scanning may improve visibility and early disruption detection (Manhart et al., 2020; Bode et al., 2011). This can enhance firms' capacity to react promptly to unexpected events, improving survival and recovery times (Essuman et al., 2022). Past studies affirm the roles of information sharing, information technology, and environmental scanning activities in enhancing resilience (e.g., Essuman et al., 2022; Gu et al., 2021; Manhart et al., 2020). Therefore, the study argues that internationally diversified firms can achieve resilience through foreign market scanning.

H2: Foreign market scanning positively mediates the relationship between foreign diversification and firm resilience.

Boundary Condition Role of Supply Chain Disruption

OIP theory emphasizes that the organizational outcomes of information search and processing activities depend on contextual factors (Tushman & Nadler, 1978), particularly environmental uncertainty drivers (Srinivasan & Swink, 2018). OIP theory suggests that excessive foreign market scanning under low uncertainty may create a misfit situation, possibly rendering firms ineffective (Tushman & Nadler, 1978; Srinivasan & Swink, 2018). Supply chain disruptions introduce uncertainty as their timing and impact are unpredictable (Wong et al., 2020; Bode et al., 2011). These disruptions encompass unexpected events that interrupt the normal flow of materials and products in supply chains (Wong et al., 2020).

Firms may employ foreign market scanning to explore opportunities or mitigate threats. In high-disruption contexts, firms may concentrate foreign market scanning on gathering disruption-related information (Bode et al., 2010). Again, high-disruption environments can enable firms to experiment and gain disruption-specific knowledge, aiding the interpretation of foreign market information (Essuman et al., 2022). Therefore, such environments can help firms to deploy foreign market scanning to build resilience effectively.

Conversely, low-disruption environments can direct foreign market scanning toward opportunity exploration, with little incentive to focus on disruptions. Moreover, limited disruption-specific knowledge in low-disruption environments (Bode et al., 2010) can make increases in foreign market scanning counterproductive, as firms may be overwhelmed by information overload in such situations (Essuman et al., 2022).

Considering these dynamics, the study proposes that high supply chain disruption can amplify the capacity of foreign market scanning to transform foreign diversification into enhanced firm resilience.

> H3: Supply chain disruption positively moderates the relationship between foreign market scanning and firm resilience, such that the positive relationship between foreign diversification and firm resilience through foreign market scanning strengthens in increasing supply chain disruption scenarios.

METHODOLOGY

Research Design and Data

We aimed to test the proposed conceptual model rather than establish empirical generalizations. We collected primary data from SME exporters in Ghana. SMEs rely on export as their primary mode of internationalization (Hilmersson & Johanson, 2016). Ghana's inclusive growth-based interventions have supported international SME activities, making it a suitable context for testing our model (International Trade Centre, 2016). The sample consisted of 450 SME exporters from the Ghana Export Promotion Authority's database, employing 5 to 200 full-time workers (Adomako et al., 2022).

We used a three-year window to measure foreign diversification, foreign market scanning, and supply chain disruption. A two-year window, however, was used to measure firm resilience as it may take time to manifest (Wieland & Durach, 2021). Our study focused on cases with at least three years of export experience involving independent SME exporters registered with the Ghana Export Promotion Authority (Adomako et al., 2022).

Due to limitations in obtaining relevant and valid secondary data, most firm and supply chain resilience studies rely on primary data from cross-sectional surveys (e.g., Essuman et al., 2023; Gu et al., 2021; Yang et al., 2021). Therefore, we adopted a cross-sectional design to capture the study's variables, given the lack of suitable secondary data from developing economy SME exporters (Kull et al., 2018).

Table 1. Sample Characteristics

Characteristics	Frequency	%	Min	Max	Mean	SD
Foreign market destination of the firm[1]:						
Sub-Saharan African market	52	11				
Asia market	62	13.1				
European market (including the UK)	152	32.1				
Middle East market	131	27.7				
Central America and the Caribbean market	4	0.8				
South America market	7	1.5				
North America market	20	4.2				
Eastern European market (excluding all EU countries)	45	9.5				
Primary products that the firm deals in (industry):						
Agricultural/food products	234	86				
Toiletries	4	1.4				
Pharmaceutical products	3	1.1				
Minerals and metals	10	3.6				
Rubber and plastics	2	0.7				
Wood and furniture	16	5.7				
Chemicals	10	3.6				
Nature of products the firm deals in (value addition) [1]:						
Unprocessed products	149	36.7				
Semi-processed products	189	46.6				
Finished products	68	16.7				
Foreign diversification (number of foreign markets that the firms operate in)			1	8	3.14	1.078
Firm size (number of full-time employees)			5	152	22.17	24.407
Firm age (number of years in operation)			3	37	15.32	7.086
Firm's international experience (number of years engaged in international business activities)			3	30	12.76	6.125

Note: [1]Multiple response.

To enhance the response rate, we employed a face-to-face approach to collect data (Essuman et al., 2022; Adomako et al., 2022). Of the 391 firms approached, 375 consented to participate, and 272 completed questionnaires were retained for analysis, resulting in a 69.57% effective response rate. The sample characteristics, including international experience, firm size, and export destinations, are presented in Table 1. No significant differences were found between early and late respondents' data (questionnaires received within 15 days and the following 15 days, respectively).

Considering that our sample comprises SMEs (Kull et al., 2018), we obtained data from a key informant per firm (cf., Gu et al., 2021; Yang et al., 2021; Wong et al., 2020). The informants were senior managers with an average of 9.62 years of experience in their current position. As many as 78% hold CEO or other top executive positions, and 90.1% hold at least a bachelor's degree. In line with previous studies (e.g., Essuman et al., 2020), we adapted three items anchored on a seven-point scale (1=strongly disagree; 7=strongly agree) to evaluate the informant's competence level: knowledge about the issues covered in the survey (mean = 6.23, standard deviation = 0.750), confidence in the responses provided (mean = 6.19, standard deviation = 0.765), and confident that the responses reflect organizational situation (mean = 6.11, standard deviation = 0.739). A t-test revealed that the average competence scale is significantly greater than 5.00. These results, together with the informants' position, high educational background, and high positional experience, contribute to the reliability and validity of the data.

Measurement Indicators and Questionnaire Development

To improve the reliability and validity of the data while minimizing common method bias, we drew on extant literature and feedback from expert assessments and a pilot study to identify indicators for the constructs. Table 2 presents the multi-indicators and their psychometric properties.

Foreign diversification was measured as the number of foreign markets a firm operates in (Fariborzi et al., 2022; Arte & Larimo, 2022). Firms were asked to indicate the number of foreign markets they regularly exported products to in the past three years.

To measure *foreign market scanning*, we used four items based on insights from environmental scanning literature (Yu et al., 2019) and international attention research (Bouquet et al., 2009). Firms were asked to rate the extent to which they focused on gathering data about new market opportunities, emerging issues threatening business success, resources for exploiting market opportunities, and resources for mitigating threats in foreign markets. Each item was anchored on a seven-point scale: 1 = "to no extent" to 7 = "to the greatest extent".

Firm resilience was operationalized as the ability to absorb, recover from, adapt, or transform business processes in the face of supply chain disruptions, aligning with recent conceptual works and empirical studies (Wieland & Durach, 2021; Wong et al., 2020; Gu et al., 2021). Four items reflecting the core properties of resilient firms were used: the ability to maintain structure and normal functioning of business operations amid supply chain disruptions; the ability to restore operations quickly following supply chain disruptions; the ability to adapt business processes to new, desirable states in the face of supply chain disruptions; the ability to transform existing business models quickly during supply chain disruptions. Each item was anchored on a seven-point scale ranging from "not at all (=1)" to "to the largest extent (=7)".

We used four reflective items to capture *supply chain disruption*, focusing on unexpected events that interrupt supply chain operations in foreign, cross-border, and local markets (Wong et al., 2020). The firms rated the items on a seven-point scale: 1 = not at all; 7 = to the largest extent =7.

The study controlled for firm-specific characteristics and external environmental factors that may influence foreign market scanning and firm resilience (Manhart et al., 2020). These variables included international experience (i.e., the natural logarithm of how long in years a firm has operated in foreign markets), foreign market unit (dummy variable: 1 = yes, 0 = no), firm size (i.e., the natural logarithm of the number of full-time employees), financial slack (i.e., the extent of discretionary financial resources), relational slack (i.e., the extent to which a firm works with alternative foreign business partners), geographical market type (dummy variables: European markets =1, otherwise =0; Middle East markets =1, otherwise = 0), and product characteristics (i.e., low-value-added products = 1, or high-value-added = 0).

Table 2. Results of Confirmatory Factor Analysis

Construct/indicator/congeneric reliability (ρ_C)/average variance extracted (AVE)	Min	Max	Mean	SD	Standardized loadings (t-value)
Supply chain disruption intensity (ρ_C = 0.863; AVE = 0.612). *Since January 2020,*					
we have encountered several unexpected events that interrupted our foreign market operations.	1	7	4.98	1.381	0.725(20.829)
supply chain disruptions have become very rampant in our foreign markets.	1	7	4.75	1.550	0.780(25.313)
our local supply chains have experienced frequent disruptive events.	1	7	3.96	1.654	0.793(26.951)
our cross-border business operations have been interfered with by several unexpected disruptions.	1	7	4.46	1.550	0.829(30.539)
Firm resilience (ρ_C = 0.856; AVE = 0.598). *To what extent was your company able to demonstrate each of the following characteristics when it experienced supply chain disruptions (i.e., unexpected events that interrupt product and service flows in a supply chain) in the last two years?*					
quickly returning business operations to a normal state.	1	7	4.93	1.196	0.718(20.201)
maintaining the desired level of control over the structure and function of core business activities.	1	7	5.03	1.168	0.771(24.308)
quickly adapting business processes to new, desirable states.	1	7	4.93	1.110	0.829(29.979)
transforming existing business models rapidly.	1	7	4.80	1.133	0.771(24.524)
Foreign market scanning (ρ_C = 0.822; AVE = 0.536). *Considering your company's operations in foreign markets (outside Ghana) within the last three years, kindly indicate the extent to which your company has focused significant time and effort on gathering data about…*					
new market opportunities in foreign markets.	1	7	4.36	1.381	0.751(20.515)
emerging issues that threaten business success in foreign markets.	1	7	4.22	1.435	0.730(19.296)
resources (e.g., technologies, processes) for exploiting new market opportunities in foreign markets.	1	7	3.82	1.398	0.744(19.928)
resources (e.g., technologies, processes) for mitigating threats in foreign markets.	1	7	3.78	1.321	0.702(17.680)

Notes: Model fit indices: Chi-square (χ^2) = 64.247, degree of freedom (DF) = 51, p = 0.101, normed χ^2 = 1.260, root mean square error of approximation (RMSEA) = 0.031, comparative fit index (CFI) = 0.990, non-normed fit index (NNFI) = 0.987, standardized root mean square residual (SRMR) = 0.036.

DATA ANALYSIS AND RESULTS

We used covariance-based confirmatory factor analysis (CFA) in Mplus to validate the reflective indicators for firm resilience, foreign market scanning, and supply chain disruption intensity. We examined and controlled for common method bias using chi-square difference test procedures in CFA and the marker variable approach. All results indicate that CMB is unlikely to confound the study results.

Key Findings

Table 3 shows the descriptive statistics and correlations for the study's variables. We utilized PROCESS 3.5 in SPSS to test our direct (H1), indirect (H2), and conditional indirect (H3) effect hypotheses. PROCESS allowed us to test the indirect and conditional indirect relationships using bootstrapping procedures. As shown in Table 4, the study finds that foreign diversification has an insignificant relationship with firm resilience: $\beta = 0.025$, $SE = 0.062$, $p = 0.683$; thus, the data do not support H1a or H1b. However, the results show that foreign diversification has a significant positive relationship with foreign market scanning ($\beta = 0.187$, $SE = 0.069$, $p = 0.007$), and foreign market scanning, in turn, has a significant positive relationship with firm resilience ($\beta = 0.176$, $SE = 0.055$, $p = 0.002$). Moreover, a bootstrapping analysis with 5,000 samples shows that foreign diversification has a significant positive indirect relationship with firm resilience through foreign market scanning: indirect $\beta = 0.033$, 95% CI [0.007, 0.069]. These results support H2.

Additional results shown in Table 5 reveal that supply chain disruption has a significant positive moderating effect on the foreign market scanning-firm resilience relationship ($\beta = 0.105$, $SE = 0.040$, $p = 0.010$). As graphed in Figure 2, a simple slope analysis reveals that foreign market scanning has a stronger and significant positive relationship with firm resilience at +1 standard deviation of supply chain disruption ($\beta = 0.293$, $SE = 0.070$, $p < 0.001$). However, foreign market scanning has a weaker and insignificant relationship with firm resilience at -1 standard deviation of supply chain disruption ($\beta = 0.022$, $SE = 0.080$, $p = 0.784$).

Using the index of moderated mediation associated with the above results, we find that the indirect effect of foreign diversification on firm resilience through foreign market scanning has a significant positive association with supply chain disruption: index of moderated mediation = 0.020, 95% CI [0.006, 0.045]. We probed these results further by analyzing the indirect effect at varying levels of supply chain disruption (Hayes, 2018). The results reveal that the indirect relationship is positive and stronger at +1 standard deviation of supply chain disruption (indirect $\beta = 0.055$, 95% CI [0.014, 0.107]) but insignificant at -1 standard deviation of supply chain disruption (indirect $\beta = 0.004$, 95% CI [-0.033, 0.044]). These results support H3.

Table 3. Correlations and Descriptive Statistics

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Firm resilience	1														
2. Supply chain disruption intensity	-0.153*	1													
3. Foreign market scanning	0.196**	-0.033	1												
4. Foreign diversification	0.059	-0.117	0.264**	1											
5. Foreign market unit	0.000	-0.076	0.235**	0.405**	1										
6. Financial slack	0.021	0.208**	-0.004	0.005	0.131*	1									
7. Relational slack	0.098	0.179**	0.098	0.156**	0.107	0.334**	1								
8. Firm size (log)	-0.009	-0.159**	0.221**	0.316**	0.516**	0.024	0.026	1							
9. Firm age (log)[a]	-0.083	-0.180**	-0.069	0.278**	0.295**	-0.113	-0.064	0.322**	1						
10. International experience (log)	-0.100	-0.219**	-0.005	0.286**	0.283**	-0.080	-0.057	0.405**	0.842**	1					
11. European market	-0.016	-0.124*	.215**	0.219**	0.115	0.011	0.050	0.139*	0.087	0.143*	1				
12. Middle East market	0.061	0.052	-0.113	0.073	0.065	-0.002	0.042	-0.078	0.069	0.016	-0.448**	1			
13. Low-value-added products	-0.039	0.074	0.025	0.029	-0.043	0.019	-0.015	0.024	0.094	0.018	-0.034	0.063	1		
14. Marker variable	-0.113	-0.080	0.027	0.040	0.014	-0.015	0.001	-0.002	-0.006	-0.002	0.029	-0.020	0.068	1	
15. Early responses	0.082	0.022	0.003	0.033	0.077	0.027	0.004	0.031	0.072	0.033	0.023	0.010	0.017	-0.009	1
Min	1	1	1	1	0	1	1	2	1	1	0	0	0	1	0
Max	7	7	7	8	1	7	7	5	4	3	1	1	1	6	1
Mean	4.92	4.54	4.05	3.14	0.33	3.92	4.35	2.73	2.61	2.42	0.56	0.48	0.55	1.89	0.53
SD	0.962	1.292	1.117	1.078	0.470	1.566	1.460	0.789	0.525	0.537	0.497	0.501	0.499	0.837	0.500
Skewness	-0.974	-0.500	-0.103	0.681	0.741	-0.080	-0.150	0.785	-0.692	-0.551	-0.238	0.074	-0.193	1.199	-0.119
Kurtosis	2.052	-0.498	-0.415	1.227	-1.462	-0.862	-0.564	0.111	0.231	-0.180	-1.958	-2.009	-1.977	1.934	-2.001

Notes: *p < 0.05(2-tailed); **p < 0.01(2-tailed); [a] = removed from the regression model due to high collinearity with international experience.

Table 4. Main and Mediation Results (H1 and H2)

	β	(B)SE	p-value	BLLCI	BULCI
HYPOTHESIZED PATHS:					
Foreign diversification (FD) → Foreign market scanning (FMS)	0.187	0.069	0.007		
FD → Firm resilience	0.025	0.062	0.683		
FMS → Firm resilience	0.176	0.055	0.002		
FD → FMS → Firm resilience	0.033	0.016		0.007	0.069
CONTROL PATHS:					
Supply chain disruption (SCD) → FMS	0.002	0.053	0.966		
SCDI → Firm resilience	-0.153	0.047	0.001		
Foreign market unit → FMS	0.307	0.169	0.071		
Foreign market unit → Firm resilience	-0.108	0.151	0.477		
Financial slack → FMS	-0.044	0.044	0.321		
Financial slack → Firm resilience	0.019	0.039	0.631		
Relational slack → FMS	0.045	0.048	0.341		
Relational slack → Firm resilience	0.064	0.042	0.134		
Firm size → FMS	0.202	0.101	0.047		
Firm size → Firm resilience	-0.013	0.090	0.884		
International experience → FMS	-0.356	0.135	0.009		
International experience → Firm resilience	-0.215	0.122	0.077		
European market → FMS	0.305	0.151	0.044		
European market → Firm resilience	-0.075	0.135	0.581		
Middle East market → FMS	-0.145	0.147	0.324		
Middle East market → Firm resilience	0.151	0.131	0.250		
Low-value-added products → FMS	0.080	0.129	0.536		
Low-value-added products → Firm resilience	-0.068	0.115	0.554		

Model fit indices:	R^2	F-value	p-value
Model of foreign market scanning	0.154	4.746	<0.001
Model of firm resilience	0.102	2.692	0.003

Notes:

1. β = unstandardized regression coefficient.
2. SE = standard error; BSE = bootstrap SE.
3. BLLCI = 95% bootstrap low confidence interval; BULCI = 95% bootstrap upper confidence interval.
4. Number of bootstrap samples = 5,000.

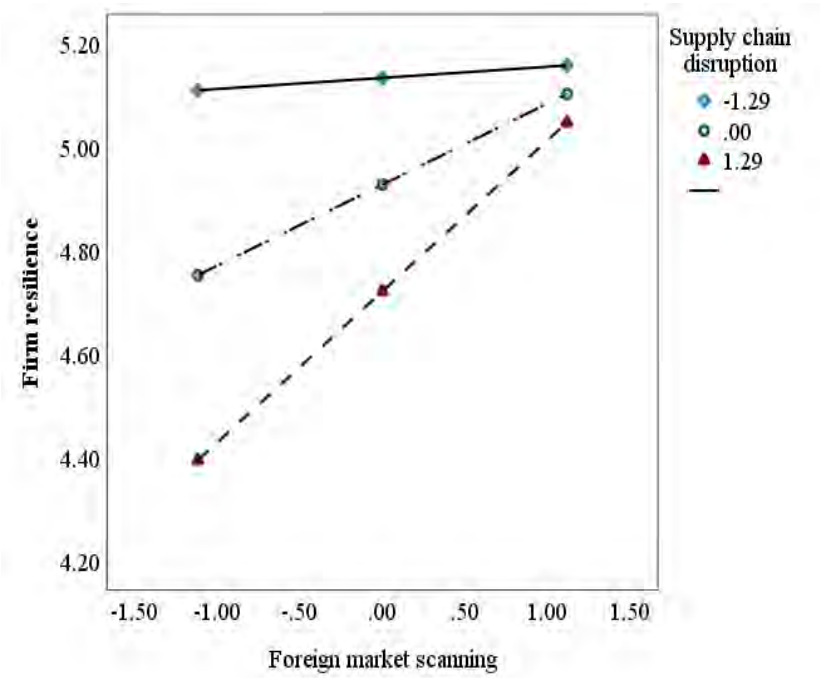
Table 5. Moderated Mediation Results (H3)

	Level of moderator	β	(B)SE	p-value	BLLCI	BULCI
HYPOTHESIZED PATHS:						
Foreign diversification (FD) → Firm resilience		0.036	0.062	0.562		
FD → Foreign market scanning (FMS)		0.187	0.069	0.007		
FMS → Firm resilience		0.157	0.055	0.005		
FMS × SCD → Firm resilience		0.105	0.040	0.010		
FMS → Firm resilience	-1SD SCD	0.022	0.080	0.784		
	Mean SCD	0.157	0.055	0.005		
	+1SD SCD	0.293	0.070	<0.001		
FD → FMS × SCD → Firm resilience		0.020	0.012		0.006	0.045
H3: IS → FMS → Firm resilience	-1SD SCD	0.004	0.019		-0.033	0.044
	Mean SCD	0.029	0.016		0.005	0.066
	+1SD SCD	0.055	0.024		0.014	0.107
CONTROL PATHS:						
Supply chain disruption (SCD) → Firm resilience		-0.158	0.046	0.001		
Foreign market unit → FMS		0.307	0.169	0.070		
Foreign market unit → Firm resilience		-0.144	0.150	0.340		
Financial slack → FMS		-0.044	0.044	0.318		
Financial slack → Firm resilience		0.014	0.039	0.724		
Relational slack → FMS		0.046	0.047	0.333		
Relational slack → Firm resilience		0.048	0.042	0.253		
Firm size → FMS		0.201	0.101	0.046		
Firm size → Firm resilience		0.004	0.090	0.967		
International experience → FMS		-0.356	0.134	0.008		
International experience → Firm resilience		-0.185	0.121	0.126		
European market → FMS		0.304	0.150	0.043		
European market → Firm resilience		-0.082	0.133	0.539		
Middle East market → FMS		-0.145	0.147	0.323		
Middle East market → Firm resilience		-0.082	0.133	0.539		
Low-value-added products → FMS		0.080	0.128	0.532		
Low-value-added products → Firm resilience		-0.047	0.114	0.679		

Model fit indices:	R^2	F-value	p-value
Model of foreign market scanning	0.154	5.293	<0.001
Model of firm resilience	0.125	3.087	<0.001

Notes:
1. β = unstandardized regression coefficient.
2. SE = standard error; BSE = bootstrap SE; p = p-value.
3. BLLCI = 95% bootstrap low confidence interval; BULCI = 95% bootstrap upper confidence interval.
4. Number of bootstrap samples = 5,000.

Figure 2. Moderating effect of supply chain disruption



Notes:

1. Level of the moderator are -1 standard deviation, mean, and +1 standard deviation.
2. Mean-centered scales of the independent and the moderating variables are reported.

Robustness Checks

Eighty-two percent of our sample are small exporters. We re-estimated the proposed model using this sub-sample. The results are consistent with the full sample results.

Our arguments about how foreign diversification affects firm resilience and foreign market scanning categorize firms based on foreign market concentration or expansion levels. Therefore, we used one-way ANOVA and multiple comparisons (Bonferroni) analysis to explore whether firm resilience and foreign market scanning vary across such groups. The results indicate that firm resilience does not differ across low, moderate, and high internationally diversified firms. However, the results indicate that the foreign market scanning is significantly greater in high foreign diversification firms, followed by moderate foreign diversification firms. These results are consistent with the PROCESS results.

The competing perspectives on the foreign diversification-firm resilience link raise questions about whether the link between these variables is nonlinear. We examined this question by estimating a quadratic regression model that regresses firm resilience on foreign diversification (FD), its mean-centered product term (FD^2), and all other variables in the study. The results show that neither FD nor FD^2 significantly affects firm resilience, suggesting that foreign diversification does not have a direct or U-shaped relationship with firm resilience.

Moreover, extant literature suggests that extreme levels of foreign market scanning can result in information overload, which might undermine organizational effectiveness (Bouquet et al., 2009). Accordingly, we estimated a quadratic regression model that regresses firm resilience on foreign market scanning (FMS), the mean-centered product term of FMS (FMS^2), and all other variables in the study. The results show that FMS has a significant positive relationship with firm resilience, but FMA^2 does not significantly relate to firm resilience.

DISCUSSION

Theoretical Contributions and Implications

While literature suggests that foreign diversification can protect against disruptions (Simchi-Levi & Haren, 2022; Kersan-Škabić, 2021; Kano & Hoon Oh, 2020), the study's results reveal a non-direct link between foreign diversification and firm resilience. This finding also is inconsistent with Puhr and Müllner's (2022) finding that diversified multinationals exhibit higher resilience.

The study's focus on international SMEs may explain this divergence in findings. For example, due to resource constraints, in the absence of uncertainty-reducing mechanisms, international SMEs may be overwhelmed by the complexities associated with foreign diversification. More importantly, because foreign diversification tends to be a double-edged sword, failing to account for specific mechanisms that trigger specific effects can mask its relationship with firm resilience (Huang et al., 2023).

Consistent with the OIP perspective, this study identifies foreign market scanning as a crucial mediator that enables internationally diversified to achieve resilience. Foreign market scanning can improve understanding and quicker responses to disruptions (Essuman et al., 2022). The study's finding that foreign market scanning enhances firm resilience strengthens past research that shows how information search and processing activities and resources benefit firm and supply chain resilience capabilities (Essuman et al., 2022; Gu et al., 2021; Lorentz et al., 2021; Manhart et al., 2020).

Furthermore, the study finds that, under low disruption conditions, less emphasis on foreign market scanning is required for resilience. In high-disruption situations, increasing scanning amplifies the resilience benefits. Accordingly, when firms leverage foreign diversification through foreign market scanning, they gain superior resilience in high supply chain disruption situations. These results are consistent with the OIP perspective that enhanced firm resilience may accrue when foreign market scanning matches information processing needs (e.g., disruptions) (Tushman & Nadler, 1978; Srinivasan & Swink, 2018). Consistent with related previous studies (e.g., Essuman et al., 2022; Lorentz et al., 2021), this study's results imply the resilience outcome of foreign market scanning and how foreign market scanning mediates the foreign diversification-firm resilience relationship are context-dependent.

These insights contribute to the literature on resilient organizations and supply chains, particularly in the international business context. By applying OIP perspective, we demonstrate how foreign diversification, foreign market scanning, and supply chain disruption combine to better explain the heterogeneity in firm resilience in the context of exporters from a developing country. Specifically, this research contributes to the limited knowledge of how and under what conditions foreign diversification affects firm resilience (Puhr & Müllner, 2022).

Implications for International SMEs

Supply chain disruptions pose significant risks to international SMEs. To increase resilience, businesses can expand across multiple foreign markets. However, this may expose them to further disruptions and complicate their response. This study reveals that international SMEs can still achieve resilience advantages by investing in foreign market scanning. Senior executives should prioritize resilience-building and foreign market scanning strategies. SMEs can navigate disruptions effectively by gathering information about foreign market opportunities, threats, and resources. Senior managers should invest in foreign market research and intelligence activities and leverage networks and the internet for up-to-date information.

While supply chain disruptions are undesirable, they can enhance firms' understanding and response capabilities. The Covid-19 pandemic demonstrated the potential for improved creativity and innovation. International SMEs can leverage disruption events to enhance the efficacy of foreign market scanning and develop resilience. Investing in foreign market scanning activities can improve resilience, especially for SMEs that face more disruptions. Senior managers should actively learn from disruptions, develop knowledge capacity, and adapt their mindset to gather, interpret, and act on foreign market information in response to disruptions.

Opportunities for Research on International Firm Resilience

The study improves our understanding of the relationship between foreign diversification and firm resilience to supply chain disruptions. However, there are some limitations to the empirical results. SMEs encounter significant internationalization challenges, which raises questions about the generalizability of our findings to other international firm contexts. Additional research is needed to test the proposed conceptual model using data from large international firms or MNCs. The different implications of supplier and customer market expansion/concentration call for more research to examine the moderation effect of foreign market type (supply-based versus customer-based expansions) on our model.

Future studies can replace our mediating variable with other knowledge-enhancing variables and examine the interaction between supply chain disruption orientation and foreign market scanning. Exploring other elements of internationalization, such as pace, scale, and pattern, and applying configuration logic to orchestrate internationalization for firm resilience would be valuable lines of inquiry.

While our measurement of firm resilience incorporates static and dynamic indicators, future studies can analyze resilience at its dimensional levels and explore resilience at the supply chain level. In addition, future studies can use secondary data and using natural experiments or longitudinal surveys to address the limitations of our research design.

References

Adomako, S., Amankwah-Amoah, J., Donbesuur, F., Ahsan, M., Danso, A., & Uddin, M. (2022). Strategic agility of SMEs in emerging economies: Antecedents, consequences and boundary conditions. *International Business Review*, 1-13.

Arte, P., & Larimo, J. (2022). Moderating influence of product diversification on the international diversification-performance relationship: a meta-analysis. Journal of Business Research, *139*, 1408-1423.

Bode, C., Wagner, S. M., Petersen, K. J., & Ellram, L. M. (2011). Understanding responses to supply chain disruptions: insights from information processing and resource dependence perspectives. *Academy of Management Journal*, *54*(4), 833–856.

Duffy, C. (2023). Intel CEO: Chip supply chains will shape geopolitics more than oil over the next 50 years. https://edition.cnn.com/2023/01/17/tech/intel-ceo-chip-manufacturing-davos/index.html (accessed on April 10, 2023).

Essuman, D., Ataburo, H., Boso, N., Anin, E. K., & Appiah, L. O. (2023). In search of operational resilience: how and when improvisation matters. *Journal of Business Logistics, 44*, 300–322.

Essuman, D., Boso, N., & Annan, J. (2020). Operational resilience, disruption, and efficiency: Conceptual and empirical analyses. *International Journal of Production Economics*, *229*, 1-11.

Essuman, D., Bruce, P. A., Ataburo, H., Asiedu-Appiah, F., & Boso, N. (2022). Linking resource slack to operational resilience: integration of resource-based and attention-based perspectives. *International Journal of Production Economics*, *254*, 1-16.

Fariborzi, H., Osiyevskyy, O., & DaSilva, C. (2022). The effect of geographic scope on growth and growth variability of SMEs. *Journal of World Business*, 57(5), 1-14.

Galbraith, J. R. (1974). Organization design: an information processing view. *Interfaces*, 4(3), 28-36.

Gu, M., Yang, L., & Huo, B. (2021). The impact of information technology usage on supply chain resilience and performance: An ambidexterous view. *International Journal of Production Economics*, 1-13.

Hilmersson, M., & Johanson, M. (2016). Speed of SME internationalization and performance. *Management International Review*, *56*, 67-94.

Huang, S., Ding, Z., Lin, X., & Zhu, Y. (2023). Post‐entry internationalization speed, learning speed, and performance: A meta‐analytic review and theory extension. *Global Strategy Journal*, 13(2), 483-516.

Iftikhar, A., Purvis, L., & Giannoccaro, I. (2021). A meta-analytical review of antecedents and outcomes of firm resilience. *Journal of Business Research*, *135*, 408-425.

International Trade Centre (2016). SME competitiveness in Ghana: alliances for action. extension://elhekieabhbkpmcefcoobjddigjcaadp/https://intracen.org/media/file/2825 (accessed on February 10, 2023).

Kano, L., & Hoon Oh, C. (2020). Global value chains in the post-COVID world: governance for reliability. *Journal of Management Studies*. https://doi.org/10.1111/joms.12626

Kersan‐Škabić, I. (2022). The COVID‐19 pandemic and the internationalization of production: A review of the literature. *Development Policy Review*. https://doi.org/10.1111/dpr.12560

Kull, T. J., Kotlar, J., & Spring, M. (2018). Small and medium enterprise research in supply chain management: The case for single‐respondent research designs. *Journal of Supply Chain Management*, *54*(1), 23-34.

Lee, H., & Chung, C. C. (2022). Go small or go home: operational exposure to violent conflicts and foreign subsidiary exit. *Journal of World Business*. https://doi.org/10.1016/j.jwb.2022.101361

Lee, J. Y., Yang, Y. S., Ghauri, P. N., & Park, B. I. (2022). The impact of social media and digital platforms experience on SME international orientation: The moderating role of COVID-19 pandemic. *Journal of International Management*, 1-21.

Li, D., Zhi, B., Schoenherr, T., & Wang, X. (2023). Developing capabilities for supply chain resilience in a post-COVID world: a machine learning-based thematic analysis. *IISE Transactions*, 1-21.

Manhart, P. S., Summers, J. K., & Blackhurst, J. V. (2020). A meta‐analytic review of supply chain risk management: assessing buffering and bridging strategies and firm performance. *Journal of Supply Chain Management*, *56*(3), 66-87.
Mondal, A., Ray, S., & Lahiri, S. (2022). Family ownership, family management, and multinationality: evidence from India. *Journal of Business Research*, *138*, 347-359.
Pitelis, C., Wang, C. L., Hughes, M., & Ambrosini, V. (2023). Dynamic capabilities and international entrepreneurship. https://www.bam.ac.uk/resource/call-for-papers-dynamic-capabilities-and-international-entrepreneurship.html (accessed on August 9, 2023).
Puhr, H., & Müllner, J. (2022). Foreign to all but fluent in many: The effect of multinationality on shock resilience. *Journal of World Business*, 1-12.
Schwens, C., Zapkau, F. B., Bierwerth, M., Isidor, R., Knight, G., & Kabst, R. (2018). International entrepreneurship: a meta-analysis on the internationalization and performance relationship. *Entrepreneurship Theory and Practice*, *42*(5), 734-768.
Simchi-Levi, D., & Haren, P. (2022). How the war in Ukraine is further disrupting global supply chains. *Harvard Business Review*. https://hbr.org/2022/03/how-the-war-in-ukraine-is-further-disrupting-global-supply-chains (accessed on April 19, 2022).
Srinivasan, R., & Swink, M. (2018). An investigation of visibility and flexibility as complements to supply chain analytics: an organizational information processing theory perspective. *Production and Operations Management*, *27*(10), 1849-1867.
Tushman, M. L., & Nadler, D. A. (1978). Information processing as an integrating concept in organizational design. *Academy of Management Review*, *3*(3), 613-624.
Wieland, A., & Durach, C. F. (2021). Two perspectives on supply chain resilience. *Journal of Business Logistics*, *42*(3), 315-322.
Wong, C.W., Lirn, T. C., Yang, C. C., & Shang, K. C. (2020). Supply chain and external conditions under which supply chain resilience pays: an organizational information processing theorization. *International Journal of Production Economics*, 1-11.
Yang, J., Xie, H., Yu, G., & Liu, M. (2021). Antecedents and consequences of supply chain risk management capabilities: an investigation in the post-coronavirus crisis. *International Journal of Production Research*, *59*(5), 1573-1585.
Yu, W., Chavez, R., Jacobs, M., Wong, C. Y., & Yuan, C. (2019). Environmental scanning, supply chain integration, responsiveness, and operational performance: an integrative framework from an organizational information processing theory perspective. *International Journal of Operations & Production Management*, *39*(5), 787-814.

DECISION SCIENCES INSTITUTE

How Do Supply Chain Capabilities Impact Food Waste?

India Allen
Georgia State University
Email: iallen13@student.gsu.edu

Madhavi Latha Nandi
Appalachian State University
Email: nandim@appstate.edu

Oyovwe Akpoigbe
Georgia State University
Email: oakpoigbe1@student.gsu.edu

Subhashish Samaddar
Georgia State University
Email: s-samaddar@gsu.edu

ABSTRACT

A critical element to managing food waste is having a robust food supply chain. This study aims to explore the impact of supply chain capabilities, as it relates to supply chain visibility and supply chain coordination on food waste at the retailing stage. A research model depicting the expected relationships is proposed based on literature review and an exploratory study on four grocery stores in the United States. The underlying themes of the aforesaid capabilities that emerged from the data are presented. The proposed model can be tested for its validity by future studies.

KEYWORDS: Food Waste, Food Supply Chain, Perishability, Supply Chain Visibility, Supply Chain Coordination

INTRODUCTION

The increasing trend in food loss and waste is a global concern and challenge (UNEP, 2021). Approximately 14 percent of the food is lost between harvest and retail, and another 17 percent is wasted in retail and by consumers (FAO, 2022). In the United States, food waste is estimated to be 30% to 40% of the food supply, corresponding to 133 billion pounds and $161 billion worth, in 2010 (FDA, 2023). Food waste is not only a societal issue in terms of not fulfilling its purpose of feeding the people, but also an environmental issue (Liu et al., 2020). Food products that end up not being consumed are associated with negative environmental externalities in the form of wasted food production processes (Kaipia et al., 2013). Such negative externalities of food production include ecotoxicity from pesticides, eutrophication, soil erosion, and biodiversity loss due to land use (Beretta et al., 2013).

The objective of this study is to assess the influence of supply chain capabilities in reducing food waste at the retailing stage of supply chain. Research on food supply chains so far has predominantly focused on identifying the root causes of food waste across different stages of supply chain (Chabada et al., 2014; Mena et al., 2011), quantification of food waste (Cicatiello, n.d., 2016; Parfitt et al., 2010), designing traceability systems and logistic solutions (Pal & Kant, 2019; Patidar et al., 2021; Hong et al., 2011), and developing inventory routing models (Chaudhary et al., 2018; Shaabani, 2022). Among the studies that investigated the root causes of food waste in supply chain, Mena et al. (2011) in their exploratory study on food supply chains in UK and Spain identified nine management related root causes for the food waste. The dominant root causes in this study specific to supply chain were lack of information sharing, lack of coordination for forecasting, ordering and shelf-life management, and cold chain management. The study by Kaipia et al. (2013) on three fresh food supply chains in Nordic countries, reiterated the importance of information sharing among supply chain partners in effectively managing the lead time of the fresh food products. Similarly, Chabada et al. (2014) in their study on six Norwegian companies of a food supply chain echoed the importance of capturing information about current product volumes and their expiration dates, appropriate forecasting, and safety stock management, to effectively manage the shelf-life and mitigate the food waste at the retail stage of the supply chain.

Resource based theory (RBT), a dominant theoretical paradigm in strategic management, posits that firms can achieve competitive advantage through acquisition, accumulation, and control over their bundles of unique resources and capabilities (Bharadwaj, 2000; Nandi et al., 2020). A firm's resources include both tangible assets (e.g., information technology infrastructure) and intangible assets (e.g., business process knowledge) possessed by the firm that enable the production and delivery of goods and services (Bharadwaj, 2000). Capabilities refer to a firm's approach to deploy and utilize the assets it possesses into the business processes to produce the desired outcomes (Grant, 1991). Supply chain capabilities refer to the ability of an organization to identify, utilize, and assimilate both internal and external resources and information to facilitate the entire supply chain activities (Wu et al., 2006). The capability of creating improved visibility at external and internal linkages in a supply chain is critical to improving supply chain performance (Barratt & Oke, 2007). While supply chain visibility is an informational capability, supply chain coordination corresponds to an operational capability (Somapa et al., 2018). Supply chain coordination represents the ability of firms to manage the dependencies within the internal functional units and external supply chain partners as they organize and perform the business activities (Arshinder et al., 2008; Nandi et al., 2020).

The capabilities of supply chain visibility and supply chain coordination and their impact on supply chain performance are extensively discussed in the general supply chain management literature. However, these capabilities are under-explored in the context of food supply chains. These capabilities are expected to assume more relevance to food supply chains given the complexities of managing food products including stringent food safety compliances, variable demands, required storage and transportation conditions to ensure product freshness (Sharma et al., 2022). Our study aims to contribute to this gap in knowledge by investigating the research question - What is the impact of supply chain capabilities – supply chain visibility and supply chain coordination on food waste?

In this paper, we present the research propositions and the research model based on the review of literature. Since the research that motivated this study is predominantly conducted in Europe, we conducted an exploratory study to validate the relevance of the proposed model to the US context. The proposed research model is later modified incorporating the findings of the exploratory study. The paper is organized as follows: Section 2 covers the review of literature relevant to the study's objectives and research propositions. Section 3 discusses the methodology used to achieve the research objectives of the study. Section 4 presents the results of the study. The paper ends with Section 5 offering some concluding remarks and limitations of the study.

LITERATURE REVIEW

Food Waste

Food waste has been defined in different ways in industry and research. The Food and Agriculture Corporation defined food waste as the wholesome edible material intended for human consumption, arising at any point in the food supply chain that is instead discarded, lost, degraded, or consumed by pests (FAO, 1981). The World Resources Institute defined food waste as the food that is of good quality and fit for human consumption but that does not get consumed because it is discarded either before or after it spoils (Lipinski et al., 2013). Gustavsson et al. (2011) in their presentation at the Save Food Congress defined food waste as the finished food products that are manufactured, wholesaled, and retailed but not sold to the intended customer or which are sold but to a price lower than the intended price due to insufficient quality or short remaining shelf life. For this study, we adopt the FAO (1981) definition of food waste as wholesome edible material intended for human consumption that is discarded, lost, degraded, or consumed by pests during the associated distribution and transportation activities. Food is lost or wasted throughout the food supply chain, from the initial stage of agriculture to the final consumption stage (Papargyropoulou et al., 2014). The method of measuring the quantity of food post-harvest is usually by weight, although other units of measure include calorific value, quantification of greenhouse gas impacts, lost inputs (e.g., nutrients and water), and economic worth of the lost food (Parfitt et al., 2010). A major portion of the food products consumed in the US are sourced from other locations and 30% of the food entering the grocery stores is wasted without reaching the end customers (Recycle Track Systems, 2021). Managing food waste has been acknowledged as a key differentiating factor for cost savings and sustainability performance among the grocery stores (McKinsey & Company, 2022).

Supply Chain Coordination

Modern supply chains are characterized by numerous value-adding activities spanning a long chain of organizations and functions over a lengthy time horizon (Arshinder et al., 2008). Effective coordination of these activities among these organizations requires an explicit definition of processes, responsibilities and structures aligned with the overall supply chain objectives (Arshinder et al., 2008). Supply chain coordination capabilities refer to the set of capabilities that enable the firms to manage the interdependencies within their internal functions and with their external supply chain firms and afford joint efforts to work towards mutually defined goals (Arshinder et al., 2008). Firms in the supply chain need to coordinate the information flows, material flows and financial flows to ensure smooth functioning of the operations and accomplish the overall system objectives (Handayati et al., 2015; Sahin & Robinson, 2002). Effective coordination is critical for food supply chains due to the larger scale of interdependencies for value creation. Further, the food products flowing through the supply chain need to be factored for their seasonality, perishability, safety, and traceability (Handayati et al., 2015).

Supply chain coordination can be achieved by using various mechanisms. The notable mechanisms that are documented in the literature include supply chain contracts, information technology and information sharing, joint decision making, and collective learning (Arshinder et al., 2008; Handayati et al., 2015). For this study, the coordination mechanism of joint decision making is considered as an indicator for supply chain coordination capabilities. Coherent decision making helps in handling complexity and exceptions, resolving conflicts among the supply chain members and minimizing operational costs (Arshinder et al., 2008). Joint decision making is an apt choice for the study since the outcome of food waste is largely dependent on managing the uncertainties and stringent quality conditions associated with the food products as they move along the supply chain to the end customer. The decision made by each actor significantly affects other actors' decisions related to production and replenishment processes. For example, if a retailer in the supply chain decides to run a promotional event on a food product during the festival season, the distributor or supplier needs to ensure that there is a surplus of those food products available at the required time with sufficient shelf life, and that the other food chains are not affected by the promotional activity. Joint decision making therefore, is an important coordination mechanism that can be used to manage such instances. The mechanism of information technology and information sharing is also critical for managing food waste. However, it needs to be noted that even under conditions of full information availability, the performance of the supply chain can be sub-optimal if the decision makers do not use the available information for achieving overall supply chain objectives (Sahin & Robinson, 2002). In this background, information technology and information sharing, as a part of supply chain visibility is considered as a precedent but not a sufficient condition for managing food waste. Rather, supply chain coordination enabled by supply chain visibility is negatively associated with food waste in the supply chain.

> *Proposition (P3 in Figure 1): Supply chain coordination is negatively associated with food waste.*

Supply Chain Visibility

Over the past two decades, academicians and practitioners discussed supply chain visibility using different perspectives. While some discussed supply chain visibility primarily as an attribute of software or information technology, some related it to the information-sharing among the supply chain actors, and some viewed it from a process perspective by including the action and decision-making aspects (Francis, 2008; Barratt & Oke, 2007). The most accepted definition of supply chain visibility is by Barratt and Oke (2007, p.1218) that is "the extent to which actors within a supply chain have access to or share information which they consider as key or useful to their operations and which they consider will be of mutual benefit". The supply chain visibility can be achieved by the supply chain firms collecting the information from the downstream and upstream activities and sharing the relevant information with the other firms in the supply chain. Information being collected may include the actual sales, forecasted demand, customer preferences and reactions, inventory levels, production status, logistic details, and delivery lead times (Kamble et al., 2020). The supply chain visibility capability ensures that essential information is readily available to those who need it, inside and outside to the organization, for monitoring, controlling, and adapting the supply chain operations and strategies (Kamble et al., 2020). Consequently, supply chains possessing the supply chain visibility, can coordinate effectively on the key operational decisions with their supply chain partners and mitigate the ordering and delivery discrepancies (Barratt & Oke, 2007; Brandon-Jones et al., 2014; Jüttner & Maklan, 2011).

Achieving supply chain visibility in food supply chains requires effective use of both formal and informal information exchange mechanisms among the supply chain partners. The formal information exchange mechanisms are dependent upon hard assets such as the required infrastructure for tracking and sharing the status of the products (Barrat & Oke, 2007). Supply chain visibility enabled by such infrastructure may be referred to as *technological visibility*. Technological visibility refers to the extent to which the information required to coordinate the supply chain processes is automatically captured and shared among the supply chain partners using the machine-readable automatic identification and data capture (AIDC) technologies and inter-organization information systems (Somapa et al., 2018). In addition to the availability of required information, the information also needs to be usable for the coordination timeliness terms of its accuracy, timeliness and completeness (Barratt & Barratt, 2011; Brandon-Jones et al., 2014).

Alongside the formal information exchange mechanisms, research has found a prominent role of informal information exchange mechanisms in enabling supply chain visibility (Barratt & Oke, 2007; Barratt & Barratt, 2011). The informal information exchange mechanisms enable the supply chain partners to exchange information on internal and external events that might affect the supply chain's overall performance (Barratt & Oke, 2007). Such information linkages can act as bridges across any gaps in visibility between the organizations, thereby overcoming to some extent, any "structural holes" that may inhibit streamlined operations of the supply chain (Barratt & Barratt, 2011). These informal information linkages are usually fluid, invisible and not measurable, making them unique and non-imitable. The informal exchanges, however, can be characterized by the nature of the relationship between the supply chain partners. If the supply chain partners share a good relationship in terms of high levels of trust, commitment, and loyalty, they are more likely to have high levels of information sensing and exchange. Supply chain visibility enabled by these relational linkages may be referred to as *relational visibility*.

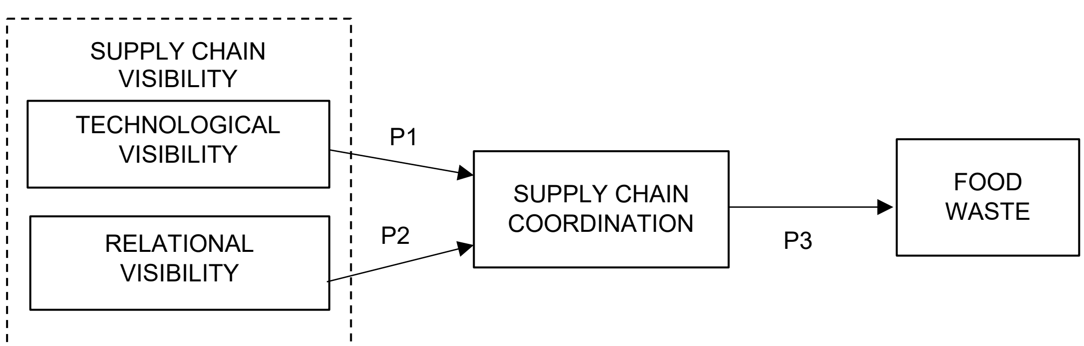
Technological and relational visibility are expected to have a positive influence on the coordination of physical movements and decision-making within the supply chain.

Proposition (P1 in Figure 1): Technological visibility is positively associated with supply chain coordination capabilities.

Proposition (P2 in Figure 1): Relational visibility is positively associated with supply chain coordination capabilities.

The proposed research model is shown in Figure 1.

Figure 1: Proposed Research Model



METHODOLOGY

A preliminary exploratory study was conducted to validate the relevance of the proposed research model in the US context. The unit of analysis for the study is the first level food supply chain network of a grocery store. The focal unit is the grocery store.

Data Collection

The exploratory study involved in-depth semi-structured interviews of four grocery stores with varied attributes. The details of the grocery stores are presented in Table 1. The interviewees are managerial level employees who are knowledgeable of their store's fresh-food management processes.

Table 1: Case Details

CASE ATTRIBUTE	GS1	GS2	GS3	GS4
Location	Large city in Georgia	Small town in North Carolina	Suburb of a large city in Illinois	Large city in California
Business Structure	Stand-alone store	Part of a supermarket chain having 300 stores in seven states.	Part of a multinational supermarket chain having 4624 stores across the US.	Part of a multinational big-box retail chain having 589 stores across the US
Food Product Range	All types of food products	All types of food products, Bakery, Deli	All types of food products, Bakery, Deli	All types of food products, Bakery, Deli
Team Size & Structure (for Food Product Management)	6 associates, 1 supervisor, 1 store manager	40 associates, 3 co-managers, 1 store manager	60 associates, 5 team leads, 1 coach, 1 store manager	40 associates, 5 supervisors, 1 DND (do not destroy) supervisor, 1 fresh manager, 1 food manager, 1 store manager
Upstream Supply Network	Suppliers	Warehouse, Logistics department, Corporate office	Warehouse, Logistics department, Home office	Warehouse, Logistics department, Corporate office

Data Analysis

A broad coding scheme comprising of the constructs being considered for the study's research model is used to code the case-wise interview content. Thereafter, coded data from the cases is consolidated into construct tables and analyzed for themes within the construct. To examine the interrelationships between the constructs, causal chains are developed and analyzed within each case for outcomes that influenced the mitigation of food waste. These causal chains are then compared with the proposed relationships of the research model. The research model is then modified to incorporate the findings from the empirical data.

RESULTS

The results of the analyses are presented in two sections. In the first section, the themes that emerged from interview data for each construct are presented. In the second section, the interrelationships between the constructs that emerged from analysis using causal chains are presented.

Constructs and Constituent Themes That Emerged from Case Data

Supply Chain Visibility

Supply chain visibility in the study's context refers to the extent to which the stores and their upstream supply chain network partners collect and share the information required for their daily operations. The data analysis revealed instances where the case organizations relied on both technological and relational visibility as presented in Table 2 and Table 3.

Technological visibility is predominantly evident in the case organizations for inventory tracking, shipment arrival and unloading information, and temperature tracking during the transit. All the grocery stores use in-store inventory tracking systems and have visibility of the current levels of product stocks. The inventory information is further used for placing new orders by GS1 and GS2. While GS1 used the current inventory data to place the new orders, GS2 used an analytic tool that uses two years of historical data to make a forecast for the orders. In the case of GS3 and GS4, the orders for restocking are placed by their corporate offices. While it is expected that their corporate offices will be using the inventory data and some kind of analytic tools, precise details on the usage of such information and tools are not known. Two grocery stores, GS2 and GS3 emphasized the importance of temperature tracking systems during the transit for maintaining the food product quality as they arrived at the stores. Visibility of shipment arrival and unloading information is emphasized in GS3 and GS4 cases for organizing the required personnel for timely and effective unloading process. In the case of GS3 the visibility is achieved through an app and radio frequency identification (RFID) tags, while an inter-organizational information system (IOIS) enabled the visibility for GS4. A set schedule is followed by GS2 for the truck arrival and unloading and did not require additional information sharing interventions. GSI and GS2 highlighted the usability of the inventory information for purposes other than getting a snapshot of current inventory status. GS3 highlighted the inaccuracies that creep into inventory information due to errors of inputting wrong product codes during the checkout process. These errors trickle forward into placement of restocking orders, thereby resulting in excess product volumes and subsequent food waste. GS3 also mentioned a recent instance of waste shipment with another store, due to an error in the temperature tracking unit. Overall, the case data underlined two basic tenets of technological visibility that are *availability of relevant information* and *usability of the available information*.

Relational visibility based out of *trust* and *commitment* is evident among the case organizations. Trust, in the study's context can be defined as the grocery store's confidence in the good will of an exchange partner (Lado et al., 2008). Commitment refers to the degree of dedication to partner cooperation that is essential for relationship continuity and realization of long-term benefits (Wu & Cavusgil, 2006). Among the case organizations, GS2, GS3 and GS4 have a vertically integrated upstream supply chain network with the parent company owning the warehouses and logistic departments that supplied the materials to the stores. GS3 and GS4 were also integrated with their parent offices for order planning. These stores overall indicated high levels of relational visibility based out of trust for the quality of the products being delivered. GS1 being a small-sized stand-alone store dealing with external suppliers directly, relied on its own pick-up for highly perishable products. GS1 relied on suppliers only for delivery of food items with longer shelf life. Relational visibility based out of commitment is visible in the case organizations GS2, GS3, and GS4 for organizing the unloading routines based on the order delivery schedules.

Table 2: Themes for Technological Visibility				
THEME	GS1	GS2	GS3	GS4
Inventory Tracking	Yes	Yes	Yes	Yes
Shipment Information sharing	-No Response-	Based on set schedule • *"we have a schedule. And, what we know, it's a set schedule, like our orders have to be placed by a certain time. And then it's delivered that next day at a certain time."*	Using App and RFID tags for placement information. • *"as the truck is arriving, we'll have a notification in our (company) app, and it will tell you track number xxx is at your door so that we know that we need to go get it"* • *"So their (corporate office's) goals are to make sure that their files, our processes for the truck breakdown (are aligned) ...like they can scan an item on it (that) will tell them whether it goes to the shelf or if they need to get the product from the backroom."*	Using IOIS. • *"we use a system called the AS400 we are able to check loads, check items, check expected arrival date..generally within today's day and the next day you can check to see what products are coming in"*
Temperature Tracking	-No Response-	Yes • *"so if it's on a refrigerated truck, it has a unit on it, or frozen truck, which has a unit on it, there's a system within that keeps that regulated."*	Yes • *"meat will be altogether in climate-controlled truck. So, they have like these thick pads that separate everything off and it's all, climate controlled by each section. Meat and produce will be together."* • *"We do have to temp everything too.. dairy is generally going to be somewhere at 35 and under. Frozen can start anywhere from 10 to 0 and negative, so we do have to do the quality checks off of those for temperature."*	-No Response-

Table 2: Themes for Technological Visibility (Continued)				
THEME	GS1	GS2	GS3	GS4
<u>*Inventory Information Usability for Ordering*</u>	Ordering is done at store level using inventory information. • *“once we've sold them, the inventory is updated and tells us, okay, you have this amount of this product left on your shelf. So, we can reorder.”*	Ordering done at store level, using an analytic tool that relies on historical data of two years. • *“we have tools in our ordering system that gives suggestions on what to order through the history. “it gives you a basic forecast of what it says that you're going to need from this truck to that truck”* • *“So, it's not always just you trying to guess. There's factual historical information that will help you make those decisions. It always runs off the past two years. obviously, it might not account for inflation, so you always have to adjust depending on the time of the year.”*	Ordering done by home office based on inventory information. • *“for example, the counter showed that the zucchini was out of stock, but the system showed that the zucchini counter was in-stock. Typically, they try to check the system inventory by 9:00 AM, to ensure the on-hand levels are accurate, and replenished within the next day.”*	Ordering done by the corporate office based on inventory information. • *“we don’t really have control over what we receive, like what loads are coming in. That is above us.”*

Table 3: Themes for Relational Visibility				
THEME	GS1	GS2	GS3	GS4
<u>*Trust*</u>	Maintained visibility based out of trust by reliance on own truck for picking up highly perishable products. • *“Some of our products, our van does the pick-up. While for others, the supplier delivers it directly to the store, mostly dry and can foods like milk, eggs, yams”*	High levels of trust-based visibility due to integrated logistics department, warehouse, and corporate office. • *“So all of our products we order here, almost all of it comes from our own warehouse. ”we don't deal directly with the manufacturers..that would be done by people at our warehouse, or corporate office. but here at the store level..Thank goodness.”* • *“Our transportation is really good..they're all a part of our company.”*	High levels of trust-based visibility due to integrated logistics department, warehouse, and home office. • *“I don't know too much about outside of chain early with supply chain, but most of it will come from like our warehouse which we have.”* • *“they (home office) work with us ahead of time for orders coming in or if they want to bring in like special promo stuff, they'll partner with the supervisor over that area.”*	High levels of trust-based visibility due to integrated decision routines between external suppliers and buyers’ team at corporate office. • *“The items are sent by the buyer. There is a team of buyers in corporate that works at corporate that handle and take care of that. If I need a hot item, and I know I need more, I will actually reach out to our buyers to give them our sales our quantities and give them what I need to request more.”*

Table 3: Themes for Relational Visibility (continued)				
THEME	GS1	GS2	GS3	GS4
Commitment	-No Response-	Information visibility based out of commitment to the set routine. • *".. what we know, it's a set schedule, like our orders have to be placed by a certain time. And then it's delivered that next day at a certain time, 4am - 7am."*	Information visibility based out of commitment to the set routine. • *"they have a certain time frame in which they have to drop off their deliveries.* • *"Truckloads can show up anywhere from 10:00 PM the day before to 4:00 AM that morning."*	Information visibility based out of commitment to the set routine and service quality. • *I know they will give us an expected arrival date of the food and dry loads that are coming."* • *"There are different buyers and different suppliers. They will change them from time to time if they are not happy with them..its (about) having the best quality."*
Commitment	-No response-	Information visibility through feedback exchanged with the supply chain partners. • *"sometimes we get product that's just inferior, that somehow got through the inspection process down in the warehouse. We've gotten in raspberries before, they had mold on them since we got all those that were thrown way. we sent emails and took pictures..we do that, take pictures, send emails, before we ever throw anything away."* • *"we tell them (corporate office) if we have issues here, and then they escalate it to the manufacturers. So that's kind of how it goes."*	Information visibility through feedback exchanged with the supply chain partners. *"there was another store just this past week where their meat produced did not temp because something must have happened with the controls on the truck. So, they ended up having to reject the entire order."*	Information visibility through feedback exchanged with the supply chain partners. • *"sometimes we will communicate with the suppliers and do a PowerPoint and share with all of the warehouses and each region. They will rank us (regions) in different categories in sales. We will take picture of the different sets of merchandising items."* • *"there are constant emails sent with corporate and DTE to track products being bad or to let us know if others are having issues. Or if we want to use a different buyer and get food from a different kind of farm. We try to have the highest-level product at the best stage available."*

Supply Chain Coordination

In the study's context, the extent to which the grocery stores work jointly with their immediate supply chain network partners to manage the product flows and quality is an indicator of supply chain coordination capabilities of the supply chain network. The themes that emerged for supply chain coordination from the case data are presented in Table 4. One interesting thing to be noted with the case organizations is that three out of four grocery stores – GS2, GS3 and GS4 have integrated supply chain networks owned by their parent companies. These grocery stores coordinated closely with the warehouses, logistic departments, and corporate offices for managing the material flows entering the stores. From an operations perspective, these supply chain network entities, even though owned by the same company may still be considered as external to the grocery stores since they have different functional goals and are locationally separated. Joint efforts of grocery store, corporate office, and warehouse for order planning and delivery are evident in the cases GS2, GS3 and GS4. In these cases, the perishability of food products is managed by keeping the order quantities sufficient for a shorter demand period in the grocery store and increasing the order frequency. In this regard, the supply chain partners actively coordinated order planning and receiving, almost daily. Optimized order planning based on inventory data and shelf-life of the product is observed in GS1, as well. Joint efforts are evident in the supply chain networks of GS2, GS3 and GS4 to effectively carry out the unloading process and prevent potential wastages at this hands-off point.

Apart from joint efforts in decision making and material flows, two other themes – culling and markdowns and shelf rotation, emerged as influencers of food waste. Both the themes relate to *internal coordination* of material flows as they move from the grocery stores' back rooms to stocking shelves and then customer shopping carts. Culling and markdowns procedures are highlighted in three cases GS1, GS2 and GS4. GS1 and GS4 stores actively followed the culling procedures by inspecting the products on the shelves or in the back-room inventory and removing the bad ones from the lots. This removal of the bad products plays an important from a customer perception. A customer is unlikely to pick a bunch of otherwise good apples if one apple from the lot has a blemish or a little tear. Markdowns speed up the sales process, especially for the products that are reaching the end of their shelf-life, thus preventing physical food waste. Shelf rotation emerged as another important factor that significantly reduced the food waste in GS1, GS2, and GS3. The theme of lack of shelf rotation was mentioned in a previous study by Mena et al. (2011) in the context of UK retailers, even though it did not make it to the list of most common managerial causes of food waste found in the study. Findings of this study reinforce the importance of shelf rotation for mitigating food waste in the US retail context.

Food Product Perishability

All the case organizations emphasized the importance of taking food product perishability into consideration for the coordination activities including order planning, transportation, unloading and internal coordination. References to perishability as presented in Table 5 are made in terms of expiration dates, remaining shelf-life, maintaining proper temperature controls, and careful handling of the product. Order sizes for highly perishable products tended to be smaller with more frequent deliveries as highlighted in all the four case organizations. Highly perishable products required more attention with the in-store shelf-life management and markdowns. Overall, the case data highlighted that perishability of a food product is a necessary condition that influences the effectiveness of supply chain coordination activities in reducing food waste.

Perishability of the food products in supply chain management literature has been extensively discussed with respect to development of various inventory models and vehicle routing algorithms. Perishability is modeled as a deterministic variable for products with a fixed lifetime and as a continuous random variable for products that deteriorate at variable rates based on the products' age, inventory volumes, and storage conditions (Chaudhary et al., 2018). Perishability as a concept affecting the supply chain capabilities has not been well-discussed in literature. In the context of grocery stores, perishability as a construct needs a clear definition and delineation in terms of limiting the percentage of product volumes with lower remaining shelf life at any point of time to be able to limit the food waste.

Causal Chains

Causal chains depicting plausible causes and effects are useful to analyze the underlying relationships between the events, actions, and/or states being observed or interpreted in the qualitative data (Miles et al., 2014). To verify the relationships in the proposed research model, causal chains within each case organization are constructed and shown in Table 6. Causal chains are useful in visually representing the sequence of themes and the related mechanisms observed in case narratives (Miles et al., 2014).

The proposed relationships between supply chain visibility and supply chain coordination are supported in the case data. Links between visibility and coordination are observed in several instances. First, inventory tracking and its accuracy enabled accurate order-placement. All the case organizations made explicit mentions of their efforts to maintain the accuracy of their inventory information. In the case of GS2, the inventory information also formed the basis for the analytic tool that made use of the historical inventory data for future order planning. GS3 explicitly mentioned how they ensure to maintain the inventory information accuracy using periodic checks. GS4 mentioned how the errors creeping into the inventory system due to faulty entry of product codes affect the order planning process. The case data clearly showed the linkages between inventory information visibility and optimized order planning. Second, visibility of information on order receiving enabled quick and effective unloading of received products. Visibility of arrival time of the trucks carrying the food loads either through notifications or set schedules, enabled the grocery stores to coordinate better in terms of making the personnel available for unloading the products from the truck. The food products also needed to be placed in their designated areas based on the type of the product. The location information provided on the RFID tags in the case of GS4 enables the right placement of the products. Third, feedback on extraneous events helped in carrying out the necessary problem solving and adjustments. This pertains to relational visibility since it involves trust and commitment to the overall goals of the network. The feedback mechanisms are observed in GS2 about the quality of the products arriving at the stores, in GS3 about the violated temperature regime for the products arriving at the store, and in GS4 about store-specific demand information for making necessary amendments to inventory based ordering decisions.

The proposed relationship between the supply chain coordination and food waste is also evident in the causal chains for all the cases. Perishability of the food products is highlighted as an influencing factor for all the coordination mechanisms and their influence in mitigating the food waste. Perishability as a factor is considered in order size planning, frequency of order receiving, maintenance of temperature regimes during transportation and in-store storage, shelf rotation and shelf-life management. All these coordination mechanisms are crucial for reducing food waste. However, perishability in terms of remaining shelf life of the products is an important condition that governed the effectiveness of these coordination mechanisms in reducing the

Table 4: Themes for Supply Chain Coordination				
THEME	GS1	GS2	GS3	GS4
Optimized Order Planning	• *“if the expiration dates, (are) shorter that means we have to order a small quantity. If the expiration date is longer, we can order a whole bunch which will last for about maybe two months or so on.”* • *“ you can only order like few boxes at a time depending on your sales, but we go buy the inventory.”*	• *“So we try to order our basic needs from truck to truck. So, like tomorrow, we get a truck. So, we're really only hoping to get enough product to last to the next row. And that way, it's the freshest.”*	• *For example, the counter showed that the zucchini was out of stock, but the system showed that the zucchini counter was in-stock. Typically, they try to check the system inventory by 9:00 AM, to ensure the on-hand levels are accurate, and replenished within the next day.”*	• *“Sushi – daily loads, produce and other food products – as being ordered by the corporate office, a lot of the fresh and wet loads come in on their own truck.”*
Joint Effort for Unloading	-No response-	• *“when they're unloading the perishable trucks that refrigeration is running the whole time..when it (the pallet) comes off of our refrigerated trailer, it goes right into one of our coolers. So, it's very quick.”*	• *“usually, they branch together as a group. We like to have them try to tackle a good chunk of them in produce first (at around 4 am) and then a couple of them branch off into meat department.”* • *So, their (corporate office’s) goals are to make sure that their files, our processes for the truck breakdown (are aligned)... like they can scan an item on it and it will tell them whether it goes to the shelf or if they need to get the product from the backroom... they'll have like the locations on it. “*	• *“I print out a ‘hot sheet’ for the day which tells me all of the loads that I am gonna be receiving today, for whatever date range I can pick.”* • *“that is the beginning of the process of what we are gonna be seeing and what comes on the load... We get with our team to make a plan to get there early to explain what we really need, and what we are low of”.”*
Internal Coordination: Culling and Markdowns	• *“there's no point of displaying an item that does not look presentable on the shelf for instance when plantain turns black and very soft, we dispose them.”*	• *“(markdowns)...that's a daily process. Every morning they'll go through..do a call so we're calling the fruits that are inferior, may have bruises... some of the lettuce maybe have brown edges. They'll pull out and then they reduce it down”*	-No response-	• *“the culling is a big part of it. It has saved us a lot in damaged and destroy and our shrink dollars. Just going through it…our produce that is bruised or have one or two bad apples (we) take it to the back and try to make good batches to take it back to the floor”*

Table 4: Themes for Supply Chain Coordination (continued)				
THEME	GS1	GS2	GS3	GS4
Internal Coordination: Shelf Rotation	• *"those that don't last long. We keep them according to their type. Some in the refrigerator and some on the shelf. when putting it on the shelf, you have to rotate... So maybe we have two remaining items on the shelf, those have to be in the front, the new ones go in the back."*	• *"If they're not doing the proper things by rotation. Like in the yogurt case, if they're getting the new stuff in, and they're just putting the stock in front, that old stuff to the back. And then you go do a deep check. you're finding all this out as a product. All that goes in the garbage."*	-No response-	• *"rotation is big and key for us to eliminate the product that goes bad. This will help us limit the amount of products that goes bad."* • *"we have tags that we are using. Different color tags that we tape on there. If we are not using it (the product) we will use a sell by date, or there are not really many expiration dates, it's really the sell through date. So that way if we need it in the next day or two, we are able to see what pallet we are going to use. That is how we keep track."*

Table 5: Perishability				
	GS1	GS2	GS3	GS4
Influence of Perishability	• *"if the expiration dates, it's a (are) shorter that means we have to order a small quantity. If the expiration date is longer, we can order a whole bunch which will last for about maybe two months or so."* • *"something like dry milk those expiration dates are not that long. So we can't order that much. Something like cereal (the expiration date) is about six months.*	• *"So a perishable item is an item that has a short shelf life, and it has to be, you know, cared for, you know, obviously, more gently than a 'drag that' item, because you've got to move it quickly. It comes in, you only have a few days to sell it before it starts to deteriorate."*	• *"Cucumber or something like. You're not getting that much replenishment because we don't wanna have to unnecessarily mark down stuff, but we should mark it down if we know we're gonna hit a short life on it because we don't want to be throwing it away."* • *"So like if it's only 5 days on like a certain vegetable and we've gone past that with that being in a cooler for selling because we had a huge quantity, we're now cutting into the time frame that a customer would be able to actually have it and have a good quality to actually eat it."*	• *In foods we have foods with a shorter shelf life. We tag them with a sell by date. A time to want to sell through before putting more products out to try to sell through those. To try to eliminate DND.* • *"meat chubs (we) have in Deli, once we open that packaging, we only have about 5 to 7 days to be able to go through that chub. So, they have to make sure they process that fully, which makes them stand their inventory tracking cuz they don't wanna keep opening main chub cases and then having a bunch of open containers and then they're wasting the life on it."*

food waste. The proposed research model is modified to incorporate the themes and relationships that emerged from the case study data and is presented in Figure 2. The final set of propositions are presented in Table 7.

CONTRIBUTION

Food waste is becoming an increasingly important issue both domestically and internationally (Buzby et al., 2014). This study emphasized the importance of effective supply chain capabilities to reduce food waste at the retailing stage of the food supply chain. The analysis highlights the impact that grocery store actors, the home store corporate office, vendors, suppliers, and logistic departments have in managing food waste in supply chain coordination. A thorough literature review and a logical interpretation of the same, provided us with our starting contributions in the form of a model and propositions. We had produced three propositions in this study, which are the basis of this model in in Figure 1: *(P1) technological visibility is positively associated with supply chain coordination capabilities, and (P2) relational visibility is positively associated with supply chain coordination capabilities, and (P3) Supply chain coordination is negatively associated with food waste*. This is our first contribution to the theory.

The model in Figure 1 has been supported by case study data. Specifically, P1 is verified with the evidence in Table 2: We have empirical evidence supporting that technological visibility is positively associated with supply chain visibility. Interviewees suggest that inventory tracking, unloading information, and temperature tracking capabilities during the transit are relevant during SCV. Grocery stores rely on the availability of relevant information and usability of the available information to effectively manage store level inventory.

The model in Figure 1 has been supported by case study data. Specifically, P2 is verified with the evidence in Table 3: We have empirical evidence supporting relational visibility that is positively associated with supply chain visibility. The stories that surfaced from this study accentuate how trust and commitment influences SCV in FSC. Grocery store managers highlight the need for trust with external partners and suppliers to ensure the quality of food products delivered. Further, the interviewees indicated that high levels of trust-based visibility due to integrated logistics departments, warehouse, and the corporate office influences relational visibility.

The model in Figure 1 has been supported by case study data. Specifically, P3 is verified with the evidence in Table 4. We have empirical evidence indicating SCC is negatively associated with food waste. In discussing their findings of SCC at the retailing stage, interviewees commented on optimized order planning, joint effort for unloading, internal coordination of culling, markdowns, and shelf rotation. Based on findings, SCC is viable when actors efficiently order food products and coordinate the perishable deliveries in a timely manner. Once products are received, results illustrate grocery store employees are required to cull and mark down products that are bruised, damaged, and no longer saleable, to maintain fresh food quality standards. In addition, rotating the products on the shelf and adhering to the shelf-by-date reduces food waste.

An additional proposition P4 emerged as we were conducting the case study. We found supporting evidence *(P4) that the negative association between supply chain coordination capabilities and food waste is influenced by perishability, which is included in the modified model depicted in Figure 2. In other words, the relationship between supply chain coordination and food waste is stronger for products with high perishability than for the products with low* *perishability.* The value of identifying P4 as a contribution to research emphasizes the need for organizations to build a process for managing perishability.

While more research is needed to explore how perishability should be defined, organizing perishability coordination activities is important to food retailers during order planning, transportation, unloading, and internal coordination to maintain food operational efficiency and freshness. Thus, the outcomes from this study might provide an understanding of how perishability is influenced by food waste at the retailing stage.

The theoretical lens for this study is the RBV theory, which provides an explanation of why RBV explicitly looks for the internal sources of SCA and aims to explain why companies in the same industry might differ in performance (Kraaijenbrink et al., 2010). Since its introduction into the field of strategic management, RBV has received significant attention with regard to how companies may obtain a sustainable competitive advantage (Gueler & Schneider, 2021). How to leverage resources in creating and sustaining competitive advantage for a firm has become the central focus for many scholars that link various types of market-based and capabilities with the ultimate financial performance of a firm (Wu et al., 2006).

This research provides an explanation of supply chain capabilities in mitigating food waste at the retailing stage. In particular, the results from this study provide evidence to retailers of the role of SCV in realizing the value of technological and relational visibility. On this background, we purport that retailers understand how accurateness of information availability and usability can influence food waste outcomes. Further, our findings depict how the value of relational visibility, as it relates to trust and commitment can influence performance results. It is essential that retailers investigate the conditions under which SCC is beneficial, so that it should not result in higher SC costs and imprecise information (Arshinder et al., 2008). Coordinating the food supply chain inter-organizationally may be challenging to supply chain management. Our work suggest that the proper use of resources can help understand the benefits driving a more efficient food supply chain capabilities in such areas of informational availability, information usability, optimized order planning, joint effort for unloading, and internal coordination.

CONCLUSION

The objective of this study is to assess the influence of supply chain capabilities – supply chain visibility and supply chain coordination in reducing food waste at the retailing stage of food supply chains in the United States. Therefore, a research model is proposed based on the review of literature. This study was verified for its relevance in the US context by conducting an exploratory study on four US based grocery stores. The findings from the study reinforced the significance of the specified supply chain capabilities in reducing food waste. The contributions of this study are three-fold. First, the study informs the researchers and practitioners with empirical evidence about the important role played by supply chain visibility and supply chain coordination in reducing food waste. Second, the study highlights the constituent themes of the considered capabilities that may be used by future research studies to operationalize the constructs and test the propositions. Finally, the study highlights the importance of perishability as a construct influencing the relationship between supply chain capabilities and food waste. There is a need to develop more clarity on how perishability as a construct can be operationalized and its influence on food waste can be validated.

Table 6: Within-Case Causal Chains	
CASE	CAUSAL CHAINS
GS1	• *Inventory tracking -> Inventory visibility -> Ordering for restocking based on current inventory data -> Less food waste due to excess stock* • *Visibility through pickup by own truck for highly perishable products -> More control over maintaining conditions for product quality during transit and unloading -> Less food waste* • *Effective culling -> Less food waste*
GS2	• *Inventory tracking and usability -> Inventory visibility -> Ordering for accurate quantities for restocking based on forecasting by the analytic tool -> Less food waste due to optimal order sizes* • *Set schedule for order receiving -> Prepared unloading process -> Less food waste resulting from unloading delays* • *Temperature tracking and usability -> Food quality maintained during the transit -> Better management of remaining shelf-life -> Less food waste resulting from reduced shelf-life* • *Visibility through feedback on bad products arriving at the store -> Better coordination at the warehouse -> Better quality products arriving at the store -> Less food waste*
GS3	• *Inventory tracking and usability -> Inventory visibility -> Ordering for restocking based on current inventory data -> Less food waste due to excess stock.* • *Set schedule for order receiving -> Prepared unloading process -> Less food waste resulting from unloading delays and reduced shelf-life* • *Visibility of order receiving and placement through App and RFID tags -> Effective unloading process -> Less food waste resulting from unloading delays and errors in placement* • *Temperature tracking and usability -> Food quality maintained during the transit -> Better management of remaining shelf-life -> Less food waste resulting from reduced shelf-life* • *Visibility through feedback on failed temperature conditions -> Resolution with problems in temperature tracking -> Better quality products arriving at the store -> Less food waste*
GS4	• *Inventory tracking -> Inventory visibility -> Ordering for accurate quantities for restocking based on forecasting by the analytic tool -> Less food waste due to optimal order sizes* • *Visibility of order receiving schedule through IOIS -> Prepared unloading process -> Less food waste resulting from unloading delays* • *Visibility through integrated decision routines -> Ordering the required product quantities -> Less food waste due to optimal order sizes*

Figure 2: Modified Research Model

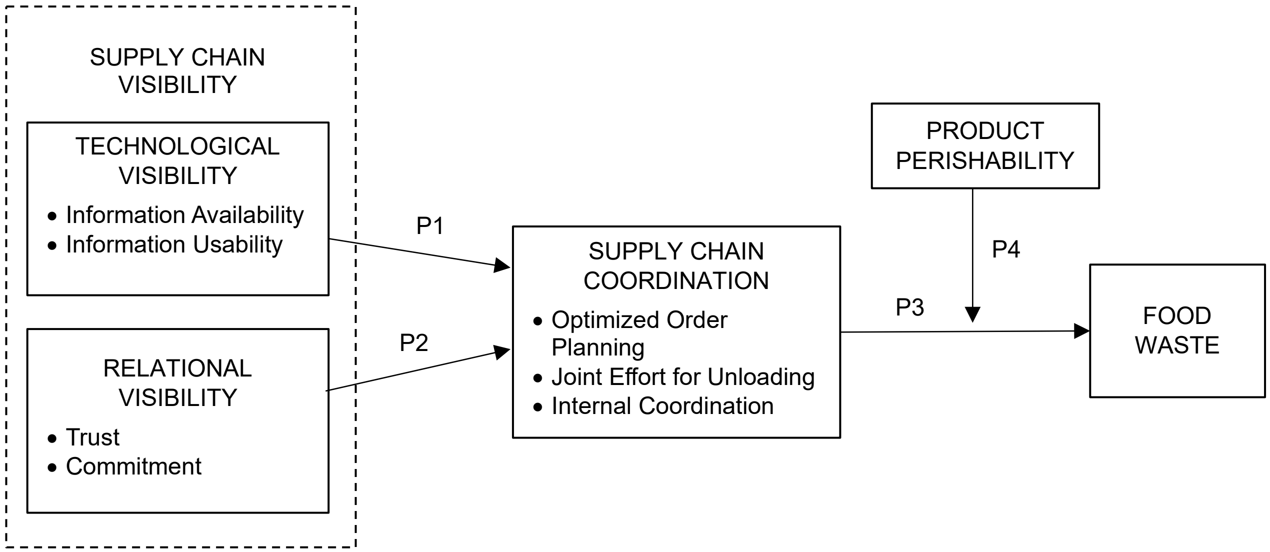


Table 7: Modified Research Propositions	
P1	*Technological visibility is positively associated with supply chain coordination capabilities.*
P2	*Relational visibility is positively associated with supply chain coordination capabilities.*
P3	*Supply chain coordination is negatively associated with food waste.*
P4	*The negative association between supply chain coordination capabilities and food waste is influenced by perishability. (In other words, the relationship between supply chain coordination and food waste is stronger for products with high perishability than for the products with low perishability).*

The study's objectives are exploratory in nature and are oriented towards informing future research studies to validate the influence of the supply chain capabilities in mitigating food waste at the retailing stage. In this regard, the findings of the study are contextual, based on the data acquired from a limited number of case organizations. Even though the case organizations are representative in terms of location, size, supply network structure and customer segments served, the findings cannot be generalized to all the grocery stores in the U.S., or elsewhere. Such a generalization may happen based on validation of the proposed research model using appropriate research designs by the future studies.

REFERENCES

Arshinder, Kanda, A., & Deshmukh, S. G. (2008). Supply chain coordination: Perspectives, empirical studies and research directions. *International Journal of Production Economics*, *115*(2), 316–335. https://doi.org/10.1016/j.ijpe.2008.05.011

Barratt, M., & Barratt, R. (2011). Exploring internal and external supply chain linkages: Evidence from the field. *Journal of Operations Management*, *29*(5), 514–528. https://doi.org/10.1016/j.jom.2010.11.006

Barratt, M., & Oke, A. (2007). Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective. *Journal of Operations Management*, *25(*6), 1217–1233.

Beretta, C., Stoessel, F., Baier, U., & Hellweg, S. (2013). Quantifying food losses and the potential for reduction in Switzerland. *Waste Management*, *33*(3), 764–773. https://doi.org/10.1016/j.wasman.2012.11.007

Bharadwaj, A. S. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*, *24*(1), 169–196. https://doi.org/10.2307/3250983

Brandon-Jones, E., Squire, B., Autry, C. W., & Petersen, K. J. (2014). A contingent resource-based perspective of supply chain resilience and robustness. *Journal of Supply Chain Management*, *50*(3), 55–73.

Chabada, L., Damgaard, C. M., Dreyer, H. C., Hvolby, H. H., & Dukovska-Popovska, I. (2014). Logistical Causes of Food Waste: A Case Study of a Norwegian Distribution Chain of Chilled Food Products. In E. Bayro-Corrochano & E. Hancock (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Vol. 8827, pp. 273–280). Springer International Publishing. https://doi.org/10.1007/978-3-662-44739-0_34

Chaudhary, V., Kulshrestha, R., & Routroy, S. (2018). State-of-the-art literature review on inventory models for perishable products. *Journal of Advances in Management Research*, *15*(3), 306–346. https://doi.org/10.1108/JAMR-09-2017-0091

Cicatiello, C., Franco, S., Pancino, B., & Blasi, E. (2016). The value of food waste: An exploratory study on retailing. Journal of Retailing and Consumer Services, 30, 96-104.

Food and Agriculture Organization of the United Nations. (1981). Food loss prevention in perishable crops. Rome: *FAO Agricultural Bulletin*, 43.

Food and Agriculture Organization of the United Nations. (2022). 2023: Tackling food loss and waste: A triple win opportunity. Retrieved from www.fao.org/newsroom/detail/FAO-UNEP-agriculture-environment-food-loss-waste-day-2022/en, May 15.

Food and Drug Administration. (2023). 2023: Food loss and waste. Retrieved from www.fda.gov/food/consumers/food-loss-and-waste, May 15.

Francis, V. (2008). Supply chain visibility: Lost in translation? *Supply Chain Management: An International Journal*, *13*(3), 180–184. https://doi.org/10.1108/13598540810871226

Grant, R.M., 1991. The resource-based theory of competitive advantage: implications for strategy formulation. *California Management Review, 33* (3), 114–135.

Gueler, M. S., & Schneider, S. (2021). The resource-based view in business ecosystems: A perspective on the determinants of a valuable resource and capability. Journal of Business Research, 133, 158–169. https://doi.org/10.1016/j.jbusres.2021.04.061

Handayati, Y., Simatupang, T. M., & Perdana, T. (2015). Agri-food supply chain coordination: The state-of-the-art and recent developments. *Logistics Research*, *8*(1), 5. https://doi.org/10.1007/s12159-015-0125-4

Hong, I.-H., Dang, J.-F., Tsai, Y.-H., Liu, C.-S., Lee, W.-T., Wang, M.-L., & Chen, P.-C. (2011). An RFID application in the food supply chain: A case study of convenience stores in Taiwan. *Journal of Food Engineering*, *106*(2), 119–126. https://doi.org/10.1016/j.jfoodeng.2011.04.014

Jüttner, U., & Maklan, S. (2011). Supply chain resilience in the global financial crisis: An empirical study. *Supply Chain Management*, *16*(4), 246–259. https://doi.org/10.1108/13598541111139062

Kaipia, R., Dukovska - Popovska, I., & Loikkanen, L. (2013). Creating sustainable fresh food supply chains through waste reduction. *International Journal of Physical Distribution & Logistics Management*, *43*(3), 262 – 276. https://doi.org/10.1108/IJPDLM-11-2011-0200

Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2020). Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *International Journal of Production Economics*, *219*, 179–194. https://doi.org/10.1016/j.ijpe.2019.05.022

Kraaijenbrink, J., Spender, J.-C., & Groen, A. J. (2010). The Resource-Based View: A Review and Assessment of Its Critiques. Journal of Management, 36(1), 349–372. https://doi.org/10.1177/0149206309350775

Lado, A.A., Dant, R.R., & Tekleab, A.G. (2008). Trust-opportunism paradox, relationism, and performance in interfirm relationships: Evidence from the retail industry. *Strategic Management Journal*, *29*, 401-423.

Liu, C., Mao, C., Bunditsakulchai, P., Sasaki, S., & Hotta, Y. (2020). Food waste in Bangkok: Current situation, trends and key challenges. *Resources, Conservation and Recycling*, *157*, 104779. https://doi.org/10.1016/j.resconrec.2020.104779

Lipinski, B., Hanson, C., Waite, R., Searchinger, T., & Lomax, J. (2013). Reducing food loss and waste.

McKinsey & Company. (2022). 2023: Navigating the market headwinds: The state of grocery retail 2022: North America. Retrieved from www.mckinsey.com/~/media/mckinsey/industries/retail/how%20we%20help%20clients/the%20state%20of%20grocery%20retail%202022%20north%20america/mck_state%20of%20grocery%20na_fullreport_v9.pdf, May 15.

Mena, C., Adenso-Diaz, B., & Yurt, O. (2011). The causes of food waste in the supplier–retailer interface: Evidences from the UK and Spain. *Resources, Conservation and Recycling*, *55*(6), 648–658. https://doi.org/10.1016/j.resconrec.2010.09.006

Miles, M.B., Huberman, A.M. & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook*. SAGE Publications.

Nandi, M. L., Nandi, S., Moya, H., & Kaynak, H. (2020). Blockchain technology-enabled supply chain systems and supply chain performance: A resource-based view. *Supply Chain Management*, *25*(6), 841–862. https://doi.org/10.1108/SCM-12-2019-0444

Pal, A., & Kant, K. (2019). Internet of Perishable Logistics: Building Smart Fresh Food Supply Chain Networks. *IEEE Access*, *7*, 17675–17695. https://doi.org/10.1109/ACCESS.2019.2894126

Parfitt, J., Barthel, M., & Macnaughton, S. (2010). Food waste within food supply chains: Quantification and potential for change to 2050. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365* (1554), 3065–3081. https://doi.org/10.1098/rstb.2010.0126

Papargyropoulou, E., Lozano, R., Steinberger, J., Wright, N., Ujang, Z. (2014). The food waste hierarchy as a framework for the management of food surplus and food waste. Journal of Cleaner Production. Volume 76. Pages 106-115. https://doi.org/10.1016/j.jclepro.2014.04.020

Patidar, A., Sharma, M., & Agrawal, R. (2021). Prioritizing drivers to creating traceability in the food supply chain. *Procedia CIRP*, *98*, 690–695. https://doi.org/10.1016/j.procir.2021.01.176

Recycle Track Systems. (2023). 2023: Food waste in America in 2023: Statistics + Facts. Retrieved from www.rts.com/resources/guides/food-waste-america, May 15.

Sahin, F., & Robinson, E. P. (2002). Flow Coordination and Information Sharing in Supply Chains: Review, Implications, and Directions for Future Research. *Decision Sciences*, *33*(4), 505–536. https://doi.org/10.1111/j.1540-5915.2002.tb01654.x

Shaabani, H. (2022). A literature review of the perishable inventory routing problem. *The Asian Journal of Shipping and Logistics*, *38*(3), 143–161. https://doi.org/10.1016/j.ajsl.2022.05.002

Somapa, S., Cools, M., & Dullaert, W. (2018). Characterizing supply chain visibility – a literature review. *The International Journal of Logistics Management*, *29*(1), 308–339. https://doi.org/10.1108/IJLM-06-2016-0150

United Nations Environment Programme. (2021). 2023: UNEP food waste index report 2021. Retrieved from www.unep.org/resources/report/unep-food-waste-index-report-2021, May 15.

Wu, F., & Cavusgil, T. (2006). Organizational learning, commitment, and joint value creation in interfirm relationships. *Journal of Business Research*, *59*, 81-89.

Wu, F., Yeniyurt, S., Kim, D., & Cavusgil, S. T. (2006). The impact of information technology on supply chain capabilities and firm performance: A resource-based view. *Industrial Marketing Management*, *35*(4), 493–504. https://doi.org/10.1016/j.indmarman.2005.05.003

DECISION SCIENCES INSTITUTE

How does digital transformation enhance supply chain performance? the role of knowledge management capabilities

Benjamin Agyei-Owusu
Kwame Nkrumah University of Science and Technology
Email: bagyei-owusu.ksb@knust.edu.gh

Francis Kofi Andoh-Baidoo
University of Texas Rio Grande Valley
Email: francis.andohbaidoo@utrgv.edu

David Asamoah
Kwame Nkrumah University of Science and Technology
Email: dasamoah.ksb@knust.edu.gh

Emmanuel Kweku Quansah
Kwame Nkrumah University of Science and Technology
Email: equansah.ksb@knust.edu.gh

ABSTRACT

Digital transformation has received increasing attention in research and practice. However, there is not much knowledge on how digital transformation enhances supply chain performance. The study proposes that digital transformation enhances supply chain performance through enhanced knowledge management capabilities. To test this assertion, empirical data was collected from 237 firms operating in a developing country and analyzed using structural equation modeling. The results indicated that digital transformation did not directly enhance supply chain performance. Rather, the effect of digital transformation on supply chain performance was fully mediated through knowledge management capabilities. The implications of the study are discussed.

KEYWORDS: Digital transformation, Knowledge Management Capabilities, Supply Chain Performance

INTRODUCTION

Digital transformation has emerged as an important phenomenon in business and management research and practice (Vial et al., 2019; Zhu et al., 2021). Digital transformation is the integration of digital technology into all areas of an organization, fundamentally changing how an organization operates and delivers value to customers (Vial et al., 2019). Whilst there is growing research on digital transformation (Chen et al., 2021; Hartley and Sawaya, 2019; Preindl et al., 2020; Agrawal et al., 2020), important gaps still remain. First, there is little knowledge on whether digital transformation enhances the performance of supply chains. This presents an important problem since it has been noted that the failure rate for digital transformation initiatives is rather high (Vial et al., 2019). To address this research gap, this study examines the effect of digital transformation on the performance of supply chains.

It is also not clear from the current literature exactly how digital transformation results in enhanced supply chain performance. Digital transformation requires close collaboration and knowledge management by members of the supply chain in order to be successful (Agrawal et al., 2020). This study therefore theorizes that knowledge management capabilities are an important knowledge-based capability that help transmit the benefits of digital transformation on supply chain performance. The assertions of the study are assessed by analyzing empirical data from 237 firms operating in a developing country.

The study contributes to literature by confirming that digital transformation results in enhanced supply chain performance of firms. Additionally, the study throws light on how exactly digital transformation enhances supply chain performance. Knowledge management capabilities are identified as key intervening mechanisms that channel the positive effect of digital transformation on supply chain performance. Specifically, the study reveals that the effect of digital transformation on supply chain performance is fully mediated through knowledge management capabilities. For practice, the study provides insights that guide managers on how they can maximize their supply chain performance through digital transformation.

The rest of the paper proceeds as follows. The theoretical background and hypotheses are discussed next, followed by a discussion of the methodology adopted for the study. Next, the results are presented and discussed. The conclusion section is finally presented.

THEORETICAL BACKGROUND

The study deploys the knowledge-based view to understand how digital transformation of supply chains creates value for firms. The knowledge-based view is an extension of the resource-based view that considers knowledge as the most important resource that can be leveraged for competitive advantage (Felin and Hesterly, 2007). The knowledge-based view, as an extension of the resource-based view, considers firms as heterogenous entities whose resource base consists of knowledge-based assets (Blome et al., 2014). A firm’s ability to create and use these knowledge-based resources can yield competitive differentiation (Kogut and Zander, 2002; Blome et al., 2014). Knowledge management plays an important role in successful digital transformation (de Bem Machado et al., 2022; Alvarenga et al., 2020). Digital transformation is less about technology and more about strategy (Kane et al., 2015). Effectively managing knowledge acquisition, conversion, application and protection is expected to be crucial to achieving positive outcomes from digital transformation initiatives (Mao et al., 2014).

RESEARCH MODEL AND HYPOTHESES

The research model of the study examines digital transformation as the independent variable and supply chain performance as the dependent variable. *Digital transformation* is conceptualized in this study as a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies (Vial, 2019). Digital transformation was conceptualized as a first order construct. *Supply chain performance* is a measure of how well the supply chain functions. Two well measures of supply chain performance are efficiency and reliability (Lee et al., 2007; Asamoah et al., 2021a). These are adopted as dimensions of supply chain performance in this study. Supply chain efficiency captures how cost- and time- efficiently the supply chain performs, whilst supply chain reliability captures the how consistently the supply chain performs. Additionally, *Knowledge Management Capabilities* is conceptualized as an intervening variable between digital transformation and supply chain performance. Knowledge management capability is an organization’s ability to obtain, generate, transfer, integrate, share and apply resources and activities that are knowledge-related across functional boundaries to create novel knowledge

(Tseng and Lee, 2014). As noted earlier, how well organizations manage knowledge is important for successful digital transformation (Agrawal et al., 2020). Four dimensions of knowledge management capabilities have been identified in the literature; knowledge acquisition, knowledge conversion, knowledge application, and knowledge protection (Gold et al., 2001; Liu and Leng, 2015). These were adopted as dimensions of knowledge management capabilities in this study. The hypotheses of the study are discussed next.

Digital Transformation and Supply Chain Performance

The knowledge-based view asserts that firms that possess high knowledge-based resources will be able to leverage these superior resources to achieve higher levels of competitive performance (Kogut and Zander, 2002; Blome et al., 2014). Firms that practice digital transformation obtain greater capabilities for sharing knowledge within their organization and with supply chain partners through digital technologies. This enables them to enjoy greater information and knowledge visibility into their operations and the operations of key supply chain partners (Asamoah et al., 2021a). This high level of visibility and knowledge sharing can enable firms to quickly vary their production volumes and variety to achieve increased supply chain performance (Asamoah et al., 2021b). Based on this, it is hypothesized that:

H1: Digital transformation is positively related to supply chain performance.

Digital Transformation and Knowledge Management Capabilities

It is anticipated that one of the benefits that digitally transformed firms obtain from their digital transformation is improved knowledge management capabilities. By incorporating digital technologies into their organizations, firms are better able to acquire, process, store and retrieve knowledge within the organization (Agrawal et al., 2020). Additionally, digital transformation typically results in the acquisition and/or upgrade of new information technology to support the digital transformation initiative. The knowledge-based resources obtained from digital technologies are typically bundled and deployed in unique ways, resulting in the creation of unique organizational capabilities for the acquisition, conversion, application and protection of organizational and supply chain knowledge (Asamoah et al., 2021a). It is therefore hypothesized that:

H2: Digital transformation is positively related to knowledge management capabilities.

Knowledge Management Capabilities and Supply Chain Performance

Achieving high levels of knowledge management capabilities is usually not the ultimate goal of firms that practice digital transformation, but rather, it is a means to an end (Mao et al., 2016). The knowledge-based view posits that firms that uniquely bundle their knowledge capabilities are able to achieve higher levels of competitive performance (Wong and Wong, 2011). This study proposes that firms that creatively deploy improved capabilities for knowledge acquisition, conversion, application and protection would be able to achieve higher levels of supply chain performance. This assertion is supported in previous studies that have noted that firms with higher knowledge management capabilities achieve higher performance (Mao et al., 2016). This leads us to hypothesize that:

H3: Knowledge management capabilities is positively related to supply chain performance.

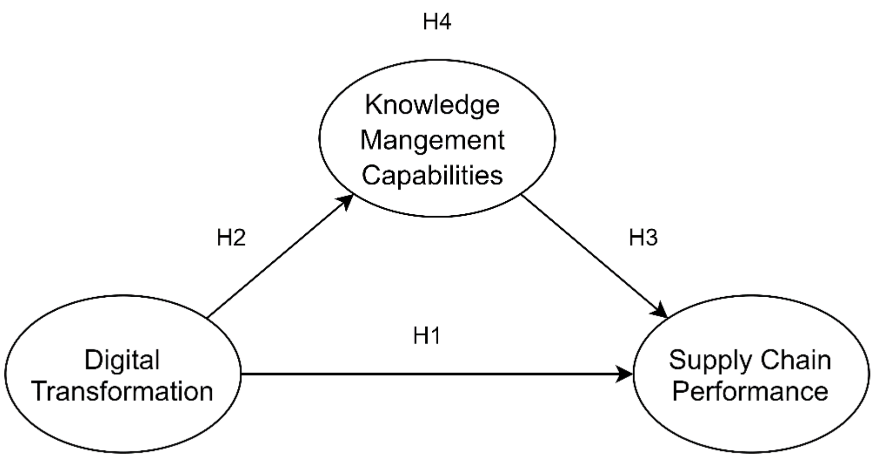
Mediating role of Knowledge Management Capabilities

Achieving improved capabilities is rarely the ultimate goal of firms that practice digital transformation. Firms that engage in digital transformation most often target clear improvements in the performance of their organizations and/or supply chains (Vial, 2019). This is what ultimately makes the significant investment of resources into digital transformation worthwhile for organizations. Whilst digital transformation is expected to result in improved supply chain performance, the mechanism through which digital transformation enhances supply chain performance is less understood. It is proposed that digitally transformed firms obtain higher knowledge-based resources, which are bundled and deployed in unique ways resulting in the creation of knowledge management capabilities. These improved knowledge management capabilities then subsequently serve as the basis for achieving high levels of supply chain performance. This leads to the hypothesis that:

> *H4: Knowledge management capabilities mediate the relationship between digital transformation and supply chain performance.*

The theoretical model of the study is presented in Figure 1.

Figure 1: Research model



METHODOLOGY

Measurement items

The measurement instrument for the constructs was adapted from previous studies. Measures for digital transformation were adapted from Nasiri et al. (2020), whilst measures for knowledge management capabilities were adapted from Liu and Deng (2015). Finally, measurement items for supply chain performance were adapted from Asamoah et al. (2021a) and Koçoğlu et al. (2011). The measures were assessed by three academics and two experts, who helped refine them to suit the context of the study. Additionally, piloting was conducted to ensure the items demonstrated sufficient validity. The final measurement items that were used in the study are presented in the Appendix.

Data collection

The empirical setting of the study is Ghana, a developing African country that has seen rapid growth in adoption of digital technologies such as enterprise resource planning systems, inter-organizational information systems, and blockchain technology (Asamoah et al., 2021a; Mintah et al., 2021). The study targeted organizations operating in Ghana's healthcare supply chain. This includes traditional healthcare providers such as hospitals and clinics, pharmacies, and medical laboratories, as well as manufacturers, wholesalers and retailers of medicines, medical and pharmaceutical products. A sample of 500 firms were randomly selected from a list of registered firms for data collection for the study. Questionnaires were sent to the firms by email, together with a cover letter detailing the purpose of the study. The targeted respondents were members of top management of the organizations. After several rounds of follow-ups, 244 responses were successfully retrieved. Seven responses had to be discarded because they had large numbers of missing responses, leaving 237 usable responses. Power analysis was conducted using a recommended medium effect size of 0.3, a minimum statistical power of 0.8, and a probability of error of 0.05 (Cohen, 2013), with the results revealing that a minimum sample size of 111 responses will be required for the results to attain statistical power. The results of the data analysis are presented next.

To test for non-response bias, the researchers compared the responses of the first 50 respondents with responses from the final 50 respondents. The statistical tests showed that there were no significant differences between the initial 50 respondents and the latest 50 respondents, suggesting that non-response bias was not a problem.

RESULTS

Demographic results

The demographic results are presented in Table 1. It can be seen that there was a fairly uniform distribution of firms based on number of employees and revenue levels. Regarding the firm ownership, privately owned firms formed the majority of respondents (54.9%).

Table 1: Demographic results		
Employees	Count	Percent
Less than 30	37	15.6
30 to 100	44	18.6
101 to 499	69	29.1
500 to 2000	55	23.2
More than 2000	30	12.7
Missing	2	0.8
Total	237	100.0
Revenue (in Ghana cedis)	Count	Percent
Less than 200,000	45	19.0
200,001 to 1,000,000	24	10.1
1 million to 5 million	62	26.2
5 million to 20 million	40	16.9
More than 20 million	60	25.3
Missing	6	2.5
Total	237	100.0

Firm Ownership	Count	Percent
State owned	80	33.8
Privately owned	130	54.9
Public-private partnership	20	8.4
Missing	7	3.0
Total	237	100.0

Measurement model results

Measurement model results were assessed by following the guidelines of Hair et al. (2019). Indicator loadings were found to be greater than 0.708 as required, indicating good item reliability. Composite reliability and Cronbach Alpha values were greater than 0.7, indicating internal consistency reliability of the research constructs. The average variance extracted (AVE) values were larger than 0.5, indicating acceptable convergent validity of the model. These results are summarized in Table 2 below.

Table 2: Psychometric properties of constructs			
Construct	Cronbach Alpha	Composite Reliability	AVE
Digital Transformation	0.887	0.914	0.640
Supply chain Efficiency	0.820	0.893	0.736
Supply chain Reliability	0.905	0.941	0.841
Knowledge Acquisition	0.834	0.889	0.668
Knowledge Application	0.758	0.892	0.805
Knowledge Conversion	0.806	0.912	0.837
Knowledge Protection	0.906	0.930	0.728

In establishing discriminant validity, we used the Fornell-Larcker criterion. Comparing the square root of the AVE of each construct with the correlation between constructs and confirming if the former is larger than the latter establishes discriminant validity (Hair et al. 2019). In Table 3 below, the diagonal values in bold which represent the square root of the AVE of constructs are larger than the off-diagonal values which are correlations among constructs, establishing discriminant validity.

Table 3: Fornell-Larcker test results							
	1	2	3	4	5	6	7
1	**0.80**						
2	0.53	**0.86**					
3	0.50	0.70	**0.92**				
4	0.55	0.56	0.59	**0.82**			
5	0.52	0.61	0.68	0.66	**0.90**		
6	0.55	0.54	0.65	0.68	0.72	**0.92**	
7	0.57	0.57	0.58	0.66	0.66	0.60	**0.85**

1 = Digital transformation, 2 = Supply chain Efficiency; 3 = Supply chain Reliability; 4 = Knowledge Acquisition; 5 = Knowledge Application; 6 = Knowledge Conversion; 7 = Knowledge Protection

The discriminant validity of the construct was also assessed using the heterotrait-monotrait (HTMT) ratio of the correlations test, which indicates that HTMT values should be less than 0.9 (Henseler et al. 2015). The HTMT values were within this range, confirming the discriminant validity of the model. The HTMT results are presented in Table 4.

Table 4: HTMT results

	1	2	3	4	5	6	7
1							
2	0.62						
3	0.56	0.82					
4	0.63	0.68	0.68				
5	0.63	0.77	0.82	0.83			
6	0.65	0.67	0.75	0.82	0.91		
7	0.63	0.66	0.64	0.76	0.79	0.70	

1 = Digital transformation, 2 = Supply chain Efficiency; 3 = Supply chain Reliability; 4 = Knowledge Acquisition; 5 = Knowledge Application; 6 = Knowledge Conversion; 7 = Knowledge Protection

Structural model results

Having established the validity of the measurement model, the structural model was assessed. The model's in-sample explanatory power was assessed by examining the R^2 of the endogenous variables. Knowledge management capabilities had an R^2 of 0.406, whilst supply chain performance had an R^2 of 0.597, which represent moderate levels of explanatory power. The predictive relevance of the model was also established since Q^2 values ranged from 0.209 to 0.725, which are significantly larger than zero (Hair et al. 2019). To ascertain whether the direct effects were supported, the path co-efficient, t-values and p-values of the hypothesized direct paths were examined. The mediation path was assessed by following the procedure outlined by Nitzl et al. (2016). The results of the hypotheses tests are summarized in Figure 2 and Table 5 below.

Figure 2: Structural model results

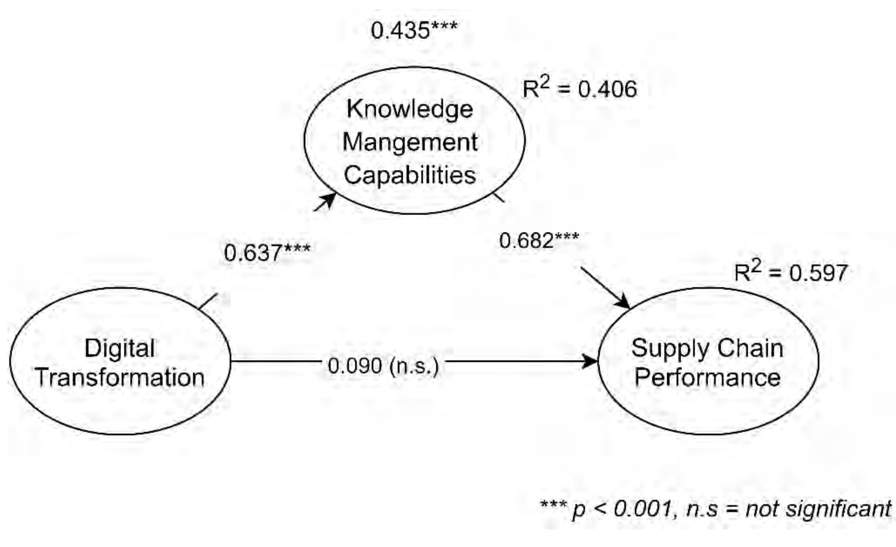


Table 5: Summary of hypotheses results

H	Hypothesis	Beta	P-Values	Support
H1	Digital Transformation → Supply Chain Performance	0.090	0.323	NO
H2	Digital Transformation → Knowledge Management Capabilities	0.637	0.000	YES
H3	Knowledge Management Capabilities → Supply Chain Performance	0.682	0.000	YES
H4	Digital Transformation → Knowledge Management Capabilities → Supply Chain Performance	0.435	0.000	YES

It can be seen from Table 5 that the direct effect of digital transformation on supply chain performance did not find significant support ($\beta = 0.090$, $p > 0.05$). This indicates that digital transformation does not directly enhance supply chain performance. It was however observed that digital transformation directly enhances knowledge management capabilities ($\beta = 0.637$, $p < 0.001$). Further, the results indicated that knowledge management capabilities significantly enhance supply chain performance ($\beta = 0.435$, $p < 0.001$), meaning hypothesis 3 was supported. Examining the indirect effect of digital transformation on supply chain performance through knowledge management capabilities revealed a positive and significant mediation effect ($\beta = 0.435$, $p < 0.001$). Given that the direct effect of digital transformation on supply chain performance was not significant, we conclude that knowledge management capabilities fully mediate the effect of digital transformation on supply chain performance. These results are discussed into further detail next.

DISCUSSIONS

Given the limited knowledge on whether and how digital transformation creates value for firms in terms of supply chain performance, the present study examined the effect of digital transformation on knowledge management capabilities and supply chain performance. It was revealed that digital transformation does not directly enhance supply chain performance. This is an interesting finding as, at first glance, it appears digital transformation does not create any supply chain performance benefits. Only through a subsequent deeper examination of the relationship between digital transformation and supply chain performance does the relationship become clearer. This finding is interesting because there have been genuine concerns about the failure rate of digital transformations (Vial et al., 2019), with some researchers indicating that significant outcomes were not observed from digital transformation (Zhu et al., 2021). Failure to examine mediation effects may result in misunderstanding the relationship between digital transformation and supply chain performance.

The study also revealed that digital transformation results in the increased knowledge management capabilities. This goes to support the assertion of the knowledge-based view that knowledge-based resources can be bundled and deployed in unique ways to create unique knowledge-based capabilities (Mao et al., 2014). In our present knowledge-intensive world, digital transformation can propel organizations into achieving high knowledge management capabilities.

The study also revealed that knowledge management capabilities directly enhance supply chain performance. This is interesting because it confirms that significant bottom-line supply chain improvements will be achieved from improved knowledge management capabilities. Organizations can seek to build their knowledge management capabilities as this translates into supply chain performance improvements for the firm.

More importantly, study sheds light on how digital transformation enhances the supply chain performance of firms. Firms that undergo digital transformation obtain knowledge-based resources, which do not directly translate into improved supply chain performance. These knowledge-based resources must be uniquely bundled and deployed, creating knowledge-based capabilities such as improved capability for knowledge acquisition, knowledge conversion, knowledge application and knowledge protection. It is these knowledge management capabilities that serve as the basis for the achievement of supply chain performance improvements.

Implications of the Study

There are some implications of the study for research. First, the study provides insights on whether digital transformation creates value for organizations. It was established that digital transformation directly enhances the knowledge management capabilities of firms and indirectly enhances supply chain performance as well. The present study therefore contributes to the emerging literature on digital transformation outcomes.

The study deployed the knowledge-based view to proffer a theoretical explanation of how digital transformation will enhance supply chain performance. Specifically, knowledge management capabilities were identified as a key variable that helps explicate the relationship between digital transformation and supply chain performance.

The study as well empirically tested the theoretical model to shed more light on the relationship between digital transformation, knowledge management capabilities and supply chain performance. The study reveals that digital transformation may not directly enhance supply chain performance outcomes, but rather indirectly enhance the supply chain performance through key intervening variables such as knowledge management capabilities.

There are some implications of the study for practice. First, the study reveals to managers that digital transformation will ultimately result in improved supply chain performance, although this may not be immediately visible. Firms can however ultimately expect improved supply chain performance from their digital transformation initiatives.

The study also identifies that knowledge management capabilities play an important role in the channeling of the full benefits of digital transformation. Managers should note that their digital transformation initiatives would first result in enhanced knowledge management capabilities, which in turn will have to be leveraged to achieve high supply chain performance.

CONCLUSION

The present study examines whether and how digital transformation results in improved supply chain performance. Improved knowledge management capabilities were identified as an important conduit that channels the effect of digital transformation on supply chain performance. Analyzing data from a survey of firms that are experiencing digital transformation revealed that digital transformation did not directly enhance supply chain performance, but rather indirectly enhanced it through improved knowledge management capabilities.

There were some limitations of the study. First, the study examined knowledge management capabilities and supply chain performance as second-order constructs. Future research can examine the proposed relationships at the first order level to create more detailed insights into the subject. The study as well relied on cross-sectional data for this study, and as such, it is difficult to claim causality between the study variables. Future research should consider using longitudinal data or secondary data to examine the relationships identified. Finally, future researchers can consider the role of supply chain learning in understanding how digital transformation enhances performance of firms and supply chains.

APPENDIX

Digital Transformation
In my organization, we aim to digitalize processes and activities as much as possible
In my organization, we collect massive volumes of data from different sources
In my organization, we aim to create stronger networking between the different business processes with digital technologies
In my organization, we aim to enhance the customer interface with digital technologies
In my organization, we aim at achieving information exchange between departments with digital technologies

Knowledge Acquisition
In my organization, we have the ability to acquire knowledge about our suppliers and customers
In my organization, we are able to generate new knowledge from existing knowledge
In my organization, we use feedback on projects to improve subsequent projects
In my organization, we are able to distribute knowledge throughout the organization
In my organization, we are able to exchange knowledge with supply chain partners
In my organization, we are able to acquire knowledge about new products/services within our industry

Knowledge Application
In my organization, we are able to apply knowledge learned from mistakes and experiences
In my organization, we are able to use knowledge to solve new problems
In my organization, we have the ability to use knowledge in the development of new products/services
In my organization, we use knowledge to improve efficiency
In my organization, we are able to quickly apply knowledge to critical competitive needs

Knowledge Conversion
In my organization, we are able to convert knowledge into the design of new products/services.
In my organization, we are able to transfer organizational knowledge to individuals
In my organization, we have the ability to absorb knowledge from individuals and service providers
In my organization, we are able to integrate different sources and types of knowledge
In my organization, we replace outdated knowledge

Knowledge Protection
In my organization, we are able to protect knowledge from inappropriate use inside and outside the organization
In my organization, we are able to protect knowledge from theft inside and outside the organization
In my organization, we have extensive policies and procedures for protecting organizational and trade secrets
In my organization, we value and protect individual’s knowledge
In my organization, we often emphasize the importance of protecting knowledge

Reliability Performance
Our firm with supply chain partners offers products that are highly reliable
Our firm with supply chain partners offers high quality products to our customers
Our firm and supply chain partners have helped each other to improve product quality
Our firm with supply chain partners increases the rate at which we fulfill customer orders
Our firm with supply chain partners increases our inventory turns

Efficiency Performance
Our firm with supply chain partners reduces inbound and outbound cost of transport
Our firm with supply chain partners reduces warehousing and inventory holding costs
Our firm with supply chain partners meets on-time delivery requirements for all product
Our firm with supply chain partners reach agreed costs per unit as compared with industry

REFERENCES

Agrawal, P., Narain, R., & Ullah, I. (2020). Analysis of barriers in implementation of digital transformation of supply chain using interpretive structural modelling approach. *Journal of Modelling in Management*, *15*(1), 297-317.

Alvarenga, A., Matos, F., Godina, R., & CO Matias, J. (2020). Digital transformation and knowledge management in the public sector. *Sustainability*, *12*(14), 5824.

Arya, B., & Lin, Z. (2007). Understanding collaboration outcomes from an extended resource-based view perspective: The roles of organizational characteristics, partner attributes, and network structures. *Journal of management*, *33*(5), 697-723.

Asamoah, D., Agyei-Owusu, B., Andoh-Baidoo, F. K., & Ayaburi, E. (2021a). Inter-organizational systems use and supply chain performance: Mediating role of supply chain management capabilities. *International journal of information management*, *58*, 102195.

Asamoah, D., Nuertey, D., Agyei-Owusu, B., & Akyeh, J. (2021). The effect of supply chain responsiveness on customer development. *The International Journal of Logistics Management*.

Blome, C., Schoenherr, T., & Eckstein, D. (2014). The impact of knowledge transfer and complexity on supply chain flexibility: A knowledge-based view. *International Journal of Production Economics*, *147*, 307-316.

Chen, L., Moretto, A., Jia, F., Caniato, F., & Xiong, Y. (2021). The role of digital transformation to empower supply chain finance: current research status and future research directions (Guest editorial). *International journal of operations & production management*, *41*(4), 277-288.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

de Bem Machado, A., Secinaro, S., Calandra, D., & Lanzalonga, F. (2022). Knowledge management and digital transformation for Industry 4.0: a structured literature review. *Knowledge Management Research & Practice*, *20*(2), 320-338.

Felin, T., & Hesterly, W. S. (2007). The knowledge-based view, nested heterogeneity, and new value creation: Philosophical considerations on the locus of knowledge. *Academy of management review*, *32*(1), 195-218.

Fornell, C. and Larcker, D. F. (1981), “Evaluating structural equation models with unobservable variables and measurement error”, *Journal of Marketing Research*, Vol 18, pp. 39-50

Hair, J. F. (2019), “When to use and how to report the results of PLS-SEM”, Vol.31 No. 1, pp. 2–24

Hartley, J. L., & Sawaya, W. J. (2019). Tortoise, not the hare: Digital transformation of supply chain business processes. *Business Horizons*, *62*(6), 707-715.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115-135.

Kane, G. C., Palmer, D., Phillips, A. N., Kiron, D., & Buckley, N. (2015). Strategy, not technology, drives digital transformation. *MIT Sloan Management Review and Deloitte University Press*, *14*(1-25).

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science*, *3*(3), 383-397.

Lee, C. W., Kwon, I. W. G., & Severance, D. (2007). Relationship between supply chain performance and degree of linkage among supplier, internal integration, and customer. *Supply chain management: an International journal*.

Liu, S., & Deng, Z. (2015). Understanding knowledge management capability in business process outsourcing: a cluster analysis. *Management Decision*.

Mao, H., Liu, S., Zhang, J., & Deng, Z. (2016). Information technology resource, knowledge management capability, and competitive advantage: The moderating role of resource commitment. *International Journal of Information Management*, *36*(6), 1062-1074.

Mintah, K., Boateng, F. G., Baako, K. T., Gaisie, E., & Otchere, G. K. (2021). Blockchain on stool land acquisition: Lessons from Ghana for strengthening land tenure security other than titling. *Land Use Policy*, *109*, 105635.

Nasiri, M., Ukko, J., Saunila, M., & Rantala, T. (2020). Managing the digital supply chain: The role of smart technologies. *Technovation*, *96*, 102121.

Preindl, R., Nikolopoulos, K., & Litsiou, K. (2020, January). Transformation strategies for the supply chain: the impact of industry 4.0 and digital transformation. *Supply Chain Forum: An International Journal* (Vol. 21, No. 1, pp. 26-34). Taylor & Francis.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The journal of strategic information systems*, *28*(2), 118-144.

Wong, W. P., & Wong, K. Y. (2011). Supply chain management, knowledge management capability, and their linkages towards firm performance. *Business Process Management Journal*.

Xu, D., Huo, B., & Sun, L. (2014). Relationships between intra-organizational resources, supply chain integration and business performance: an extended resource-based view. *Industrial Management & Data Systems*.

Zhu, X., Ge, S., & Wang, N. (2021). Digital Transformation: A Systematic Literature Review. *Computers & Industrial Engineering*, 107774

DECISION SCIENCES INSTITUTE

How does supply chain leadership relate to sustainability performance?
disaggregated and mediation analyses

Ishmael Nanaba Acquah
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
inacquah@st.knust.edu.gh / nanabaacquah8@gmail.com

David Asamoah
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
dasamoah.ksb@knust.edu.gh

Abdul-Samed Muntaka
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
abusamgh@gmail.com

Dominic Essuman
Sheffield University, Sheffied, UK
d.essuman@sheffield.ac.uk

Emmanuel Kweku Quansah
Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
equansah@hotmail.com

ABSTRACT

The study investigates how supply chain leadership (SCL) (transformational and transactional) directly and indirectly through firm resilience affects sustainability performance (social andeconomic). The hypotheses are tested on a sample of 258 Ghanaian firms using structural equation modeling. Results indicate that transformational leadership has positive relationships with economic and social performance dimensions; transactional leadership has a positiverelationship with economic performance and is unrelated to social performance; firm resiliencemediates the transformational leadership-social performance relationship and the transactionalleadership-social performance relationship. Implications for research and practice are discussed.

KEYWORDS: Supply chain leadership, Firm resilience, Sustainability performance, sub-Saharan Africa

INTRODUCTION

Contemporary research emphasizes supply chain leadership (SCL) as crucial in sustainable performance improvement (Mokhtar et al., 2019; Vivaldini and Pires, 2016; Agi and Nishant, 2017). However, this body of work has several shortcomings (Chen et al., 2021; Blome et al., 2017). First, by focusing on specific performance outcomes, previous studies say little about how SCL affects multiple sustainable performance (Blome et al., 2017; Chen et al., 2021). Second, empirical findings on the SCL-sustainability performance link are mixed and inconsistent. While some studies find SCL to be related positively to sustainability outcomes (e.g., Mokhtar et al., 2019a; Blome et al., 2017), other studies find a negative and insignificant associations (e.g., Fontoura and Coelho, 2020; Blome et al., 2017). Third, SCL is a multi-faceted construct, comprising transformational versus transactional aspects, but knowledge of which SCL dimension benefits specific sustainable performance most is lacking (Chen et al., 2021; Mokhtar et al., 2019b). Fourth, there is little theoretical explanation of the mechanisms through which SCL influences sustainability performance (Mokhtar et al., 2019; Sundram et al., 2016). Finally, there is a dearth of understanding of how SCL affects sustainability outcomes in developing economies, particularly in Sub-Saharan (Mokhtar et al., 2019; Chen et al., 2021).

In contributing to addressing the above issues, this study examines how transformational and transactional SCL affect economic and social performance in a sub-Saharan African country, Ghana. The study further proposes firm resilience as an important intervening variable that can help clarify the link between SCL and sustainable performance. The rationale is that the growing levels and costs of supply chain disruptions render the attainment of sustainable goals increasingly difficult, and it requires resilient organizations to overcome these challenges while improving sustainable performance outcomes (Negri et al., 2021). Therefore, strong SCL that aim at driving sustainable performance are inclined to engender resilience-building (Verghese et al., 2022; Mishra et al., 2021).

The study addresses two questions: *1) What are the relative effects of transformational and transactional SCL on economic and social performance? 2) Does firm resilience mediate these effects?* The study offers three key contributions. First, it decomposes the SCL constructs and examines multiple sustainable performance outcomes simultaneously, shedding detailed insights into the link between SCL and sustainability performance. Second, it identifies resilience as a significant firm mechanism that converts SCL into enhanced sustainability performance, helping to clarify the SCL-sustainability performance link. Finally, using data from an underexplored context (Ghana) broadens the understanding of the SCL-sustainability performance relationship.

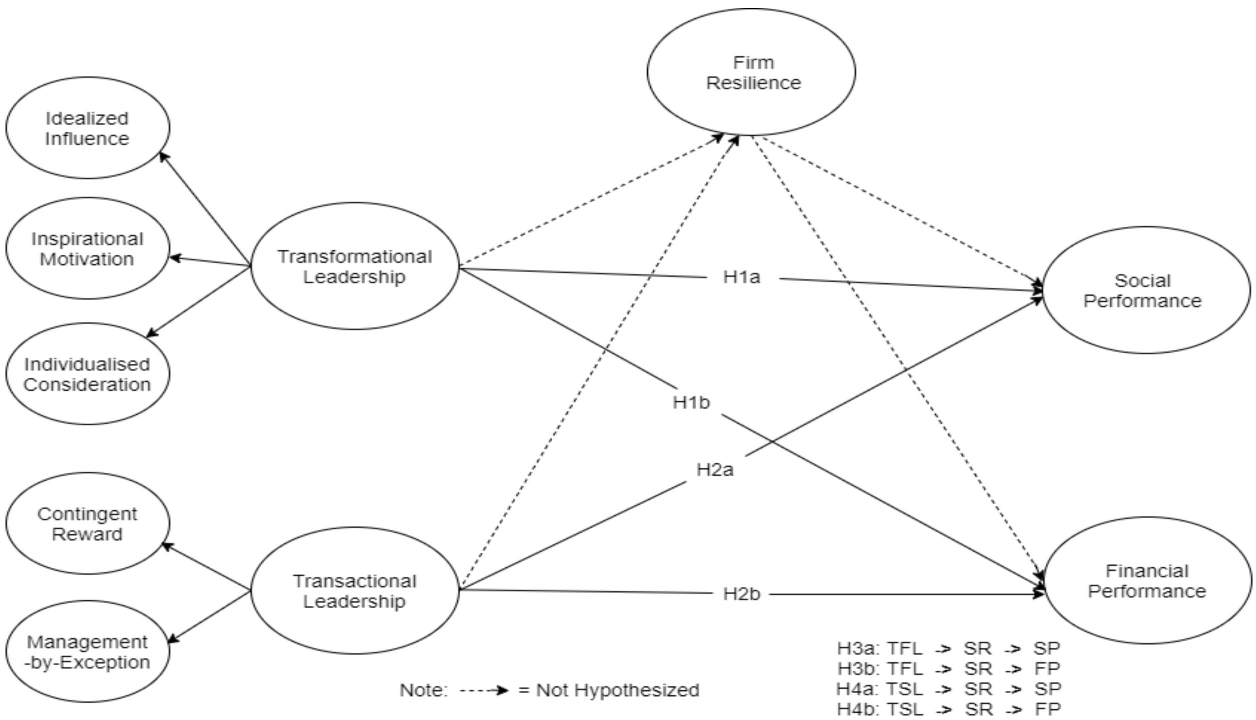
THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

The study draws on the theoretical underpinnings of the social exchange theory (SET) and dynamic capabilities theory (DCT) to examine its research questions. The SET is used to explain the relationship between SCL and sustainability performance. In contrast, the DCT is employed to explain the intervening mechanism of firm resilience in the link between these concepts.

Firm behaviour and decisions are driven by social costs and rewards (Cortez and Johnston, 2020); thus, SET crucially argues that individuals and by extension, firms enter into and maintain a relationship to obtain a net positive value (i.e., rewards>costs) (Lambe et al., 2001; Homans, 1958). A supplier that receives tangible or intangible benefits from an exchange may feel obligated to reciprocate (Terpend and Krause, 2015). Specifically, suppose suppliers in their dyadic relationship with focal firms envisage that they will obtain a net positive value from the exchange relationship. In that case, their obligation and compliance can increase, which may culminate in developing firm resilience and sustainability outcomes for parties within the supply chain.

DCT stresses the purposeful modification of resources to respond to the characteristics of an evolving business environment (Schilke et al., 2018) and, similarly, the ability of firms to build capabilities and competencies to reconfigure and continuously innovate in response to dynamic environments (Asamoah et al., 2020). The study argues from the DCT perspective that the development of firm resilience provides firm capability to mitigate and cope with upstream disruptions, thus providing the foundation for firms to respond to the supply uncertainties and complexities inherent in a present business environment to accomplish sustainable performance. Dynamic capabilities align with the resilience concept, which attempts to develop specific organizational processes and resources to facilitate sustainable competitive advantage, at least from the firm level (Dabhilkar et al., 2016; Ali et al., 2017). The study's research model is presented in Figure 1 below.

Figure 1: Conceptual model



Transformational leadership and sustainability performance

Transformational leadership in supply chain management refers to a buying firm's ability to motivate and stimulate the action and behaviours of supply chain members (Mokhtar et al., 2019a). Focal buying firms can leverage transformational leadership to drive their sustainability effort and achieve sustainable social performance. Focal firms' ability to demonstrate transformational SCL will improve the compliance and imitation of supply chain partners in line with leading firm initiatives such as corporate social responsibility (Mzembe et al., 2016). Transformational leadership is more effective in realizing superior firm performance in a present business environment characterized by dynamism and uncertainty (Lamsam and Charoensukmongkol, 2022). Transformational leaders who go beyond self-interest for the good of supply chain actors can enhance the propensity of supply chain partners towards sustainability efforts, which can result in social performance across the supply chain.

It is possible to create more sustainable businesses for every supply chain partner through a

transformational and socially responsible approach that leverages the maximum potential of relationships (Fontoura and Coelho, 2020). Transformational leadership behaviours that stress on mutual goals, aspirations and positive social exchange will translate into enhanced relationship commitment to sustainability initiatives (Birasnav et al., 2015) aimed at raising the surplus of the supply chain. Similarly, transformational or proactive leadership can be utilized by leading buying firms to enhance suppliers' full involvement and innovation towards sustainability (Mokhtar et al., 2019b) including in sustainable economic outcomes. A buying firm's transformational leadership can increase organizational innovativeness and improve the organisation's financial performance (Mokhtar et al., 2019a). Based on the above exposition, the study hypothesizes that;
H1a: Transformational leadership has a positive relationship with social performance
H1b: Transformational leadership has a positive relationship with economic performance

Transactional leadership and sustainability performance

Transactional leadership in the context of supply chain management encapsulates buying firms' behavior in clarifying suppliers' expectations and roles, rewarding in addition to monitoring and auditing of suppliers (Mokhtar et al., 2019a). Transactional leaders tend to inspect and monitor the quality of the supply process and either offer rewards or punishments based on supplier performance (Birasnav et al., 2015). Thus, the usage of transactional leadership can influence a buying firm to monitor while tracking the performance of their suppliers to ensure compliance with agreed specifications which can provide the basis for buying firms to satisfy the needs of its varied stakeholders. Buying firms can use transactional leadership to maintain the status quo concerning supply chain efforts to achieve social sustainability goals. Transactional leadership can influence suppliers to provide worker's safety, satisfaction and the development of needed skills (Wang and Dai, 2018) to meet required specification which can help buying firms in their operations and their ability to satisfy the social needs of its varied stakeholders.

Transactional leadership in establishing rules and procedures (Mokhtar et al., 2019) can contribute to supplier compliance with requirements which can facilitate the acquisition of the right quality products to facilitate production and to address the needs of varied stakeholders. This can result particularly in customer acquisition, satisfaction and retention which will inure to the economic benefit of leading buying firms. Transactional leaders will select suppliers based on cost rather than commitment which can contribute to sustainable financial performance (Birasnav et al., 2015). Thus, transactional leadership will involve selecting suppliers based on sustainability performance metrics such as cost, quality and timeliness which can crucially drive leading firm effort at achieving economic performance. Based on the above exposition, the study hypothesizes that;
H2a: Transactional leadership has a positive relationship with social performance
H2b: Transactional leadership has a positive relationship with economic performance

The mediating role of firm resilience on transformational leadership and sustainability performance

Leadership is considered to be crucial for decision-making about social practices (Fan et al., 2020; Wang et al., 2019) given that the implementation of social practices, mainly external social practices involve much costs (Liu et al., 2022). The provision of any leadership amid persistent and severe upstream disruption cannot be leveraged to influence sustainability outcomes. Firm resilience constitutes a dynamic capability (Dabhilkar et al., 2016); thus, stability in terms of supply and operations facilitated by upstream resilience from the DCT perspective can provide the foundation to utilize a transformational approach in undertaking sustainable effort to satisfy the needs of employees, customers and the external community. In dynamic environments where customers' needs change enormously and rapidly in a short period (Asamoah et al., 2021) and speed is an order winner, the development of firm resilience capabilities can provide the means to leverage transformational leadership toward addressing the needs of varied stakeholders, particularly customers.

Upstream resilience plausibly ensures continuity of material firm and stable delivery of products and is essential for firm performance and survival (Pournader et al., 2016). Amidst consistent upstream disruptions (Elliot, 2019, 2020), firm resilience enables rapid recovery while contributing to continual material and product delivery which minimizes the detrimental impacts. Transformational leadership has a positive relationship with social performance of upstream disruptions and enhances customer value and satisfaction (Chowdhury and Quaddus, 2017; Fiksel et al., 2015; Singh, 2020). Transformational leadership involving teaching and coaching suppliers can enhance a firm's profitability, sales growth and market share when there is resilience capability that results in the sustenance and continuity of supply and ensures that upstream structures and functions are guaranteed. From the DCT perspective, the ability to create and satisfy the needs of customers facilitated by upstream resilience can provide the means through which transformational leadership can be harnessed to influence sustainable economic performance. Based on the above exposition, the study hypothesizes that;
H3a: Firm resilience mediates the effect of transformational leadership on social performance
H3b: Firm resilience mediates the effect of transformational leadership on economic performance

The mediating role of firm resilience on transactional leadership and sustainability performance

Utilizing transactional leadership in maintaining definite standards to ensure sustainable social performance can materialize when leading buying firms can cope with changes brought by upstream disruption and also have the ability to adapt to upstream disruptions easily (Gu et al., 2020). Also, using lessons learnt during disruptions and unexpected events (Fan et al., 2020) can define and shape transactional engagement to ensure continual operations that can provide the basis to address the needs of customers, employees and the external community. From the DCT perspective, continuity of supply facilitated by firm resilience can help facilitate a transactional approach in driving efforts to minimize mistakes to ensure smooth operations to address the social needs of diverse firm stakeholders, resulting in sustainable social performance.

With increasing outsourcing and dependence on supply chain partners, supply chain operations cannot be sustained, and customer demands cannot be satisfied without the stability of material supply (Gu et al., 2020). Thus, the transactional approach involving the provision of rewards for achievement or punishment for mistakes (Defee et al., 2010; Mokhtar et al., 2019) aimed at accomplishing sustainable economic goals can materialize through the stability of supply which provides the basis to create value that customers appreciate. From the DCT perspective, suppliers in ensuring firm resilience will maintain situational awareness at all times during firm (Gu et al.. 2020) which can facilitate the use of transactional approach in tracking and providing full attention to upstream mistakes (Defee et al., 2010; Mokhtar et al., 2019) which can culminate in desired economic outcomes. Based on the above exposition, the study hypothesizes that;

H4a: Firm resilience mediates the effect of transactional leadership on social performance
H4b: Firm resilience mediates the effect of transactional leadership on economic performance

METHODOLOGY

A cross-sectional survey is used to investigate the study hypotheses in line with prior research. The survey's target firms were pharmaceutical firms and private healthcare facilities operating in Ghana. The study is a firm-level study, thus, respondents serving in managerial capabilities such as CEOs, managing directors and functional managers such supply chain, operations and marketing managers were the target respondents. Emerging markets have been more severely affected by the Covid-19 pandemic than developed economies in terms of organizational, operational, and economic sustainability (Harjoto and Rossi, 2021). In healthcare supply chains, supplier-related risks such as import disruptions, inefficient information-sharing means, and failure of key suppliers take precedence over operational, financial, and demand-related risks (Moktadir et al., 2018), particularly in developing regions. Also, supply chains in emerging countries are bedeviled with more barriers to sustainability which necessitates that focal firms play more significant roles in driving supply chains towards more sustainable business practices than developed economies (Silvestre, 2015).

A 7-point Likert scale was employed to measure the constructs used in the study. The Likert scale was anchored from 1 = “strongly agree” to 7 = “strongly disagree” with the midpoint anchored as “neutral”. Items used in measuring SCL (transformational and transactional) were adopted from (Mokhtar et al., 2019b). Sources for the measuring the remaining items are as follows; firm resilience (Essuman et al., 2020 and Lorentz et al., 2021), social performance (Sancha et al., 2016; Zhu et al., 2016 and Hubbard, 2009) and economic performance (Hubbard, 2009). Firm resilience was measured from the upstream or supply perspective. The measurement items were pre-tested and refined using 20 MBA students of a leading University in Ghana serving in managerial positions. A survey was conducted to collect data to analyze the study model after the survey instruments had successfully undergone pre-testing. Table 1 provides the details regarding the measurement items.

Only pharmaceutical firms and private healthcare facilities registered with the Ghana Pharmaceutical Council and Ghana Health Service were considered for the study. Additionally, only firms and facilities with more than four full-time employees were considered for the study. From a list of 1,000 firms dominantly made up of pharmaceutical firms, 350 pharmaceutical and private healthcare facilities were purposively selected for data collection. 300 of these firms were reached and willing to participate in the study. Five research assistants delivered questionnaires to these firms. In all, 278 questionnaires were retrieved, however, 20 were deemed unusable. In all, 258 useable responses were used for the data analysis, representing an effective response rate of approximately 74%. Additionally, power analysis was used to assess the sufficiency of the data. The required sample size for the study is 85, using a recommended power of 0.8, a recommended probability of error of 0.05, and a recommended medium effect size of 0.3 (Cohen, 2013). Therefore, the 258 responses used for the data analysis are adequate for a reliable study outcome. Examining non-response bias involved using the Armstrong and Overton (1977) suggested approach, specifically by contrasting early and late respondents on annual revenue and industry. The t-test results showed no statistically significant difference between the category means for firm annual revenue and industry ($p < 0.05$), indicating that non-response bias may not be a problem in this study.

Table 1: Measurement items, item loadings and psychometric properties of constructs				
Items	Measures	Loadings	CR	AVE
Regarding business relationships with suppliers, to what extent did your company demonstrate each of the following behaviors in the last three years?				
II1	Going beyond self-interest for the good of our suppliers	0.730	0.872	0.631
II2	Talking enthusiastically about what needs to be accomplished by our suppliers	0.836		
II3	Clarifying the central purpose underlying our supply chain actions	0.801		
II4	Displaying power and confidence	0.807		
IM1	Seeking different views when solving supply chain issues with suppliers	0.862	0.887	0.724
IM2	Suggesting new ways to solve supply chain issues	0.838		
IM3	Encouraging our suppliers to express ideas	0.851		
IC1	Spending time teaching and coaching suppliers	0.758	0.857	0.668
IC2	Encouraging suppliers to improve their strengths	0.859		
IC3	Giving our suppliers individual consideration	0.831		
CR1	Letting suppliers know what is expected of them in our dealings	Dropped	0.867	0.621
CR2	Deciding what shall be done and how it will be done in our exchange relationships	Dropped		
CR3	Maintaining definite standards of performance in our exchange relationships	Dropped		
CR4	Asking suppliers to follow established purchasing rules and procedures	0.711		
CR5	Rewarding suppliers for meeting performance standards	0.807		
CR6	Punishing suppliers for fault and misconduct such as late delivery	0.778		
CR7	Tracking suppliers' mistakes	0.850		
MBE1	Concentrating full attention on dealing with suppliers' mistakes	Dropped	0.849	0.653
MBE2	Believing in "if not broken, don't fix it"	0.786		
MBE3	Not interfering suppliers' production decisions and problems	0.833		
MBE4	Avoiding making decisions for our suppliers	0.805		
Indicate the extent to which your company demonstrated the following abilities when it experienced supply disruptions				
SR1	Absorbing disruption impacts	0.780	0.890	0.576
SR2	Easily recovering from disruption impacts	0.763		
SR3	Meeting normal performance objectives	0.724		
SR4	Restoring operations performance pre-disruption level over a short period	0.733		
SR5	Rapidly adjusting operating procedures to meet operational needs	0.758		

SR6	Rapidly changing operations to fulfill customer requirements	0.792		
SR7	Resuming normal operation in a cost-effective manner	Dropped		
SR8	Quickly deploying new ways to create and deliver products and services	Dropped		
To what extent has your company achieved each of the following performance outcomes in the past two years?				
SP1	Improved employee welfare	0.773	0.902	0.648
SP2	Improved occupational health and safety of employees	0.853		
SP3	Improved walfare of people in the community served	0.856		
SP4	Improved customer welfare	0.799		
SP5	Reduced customer complaints	0.736		
Over the past two years,				
FP1	Our net profit before interest on tax has been consistently above industry average	0.846	0.921	0.744
FP2	Our sales margin has been consistently above industry average	0.869		
FP3	Our return on investment (ROI) has been consistently above industry average	0.843		
FP4	Our return on assets (ROA) has been consistently above industry average	0.890		

RESULTS

The demographic data of surveyed firms were analysed using frequency tables. Regarding firm ownership, 87.21%, constituting a significant majority of the firms were locally-owned, whereas 5.81% were foreign-owned with jointly-owned firms constituting 6.98%. For the industry category, 83.72% constituting a substantial majority of the firms were pharmaceutical firms with private healthcare facilities constituting the remaining 16.28%. For the pharmaceutical type category, the highest category of firms were wholesaling with 30.10%, retailing with 28.70%, importing with 24.57% and lastly manufacturing with 16.67%. For annual revenue category, the highest category of firms constituting 35.3% had revenue between 1,000,001 and 5,000,000 cedis with the 15.9% of having the least category of firms having annual revenue above 5,000,000 cedis. The majority of the firms sampled for the study can be deemed large, as 51.2% of the firms had annual revenue exceeding 1,000,000 Ghana cedis.

Measurement model results

In analysing the data, the two-stage approach Hair et al. (2019) suggested was used. The first stage required examination of the PLS-SEM results by evaluating the measurement model, followed by the evaluation of the structural model in the second stage (Hair et al., 2019). To ascertain and ensure the quality of the measurement model, necessary tests were performed. The first stage of assessing the measurement model involved examining the loadings of the indicators (Hair et al., 2019). Items CR1, CR2, CR3 and MBE1 had loadings below the acceptable threshold of 0.708 and were therefore dropped. In addition, an assessment of internal consistency reliability using composite reliability revealed that all the items satisfy the composite reliability threshold of values above 0.7 (Hair et al., 2019). In addition, the average variance extracted (AVE) were higher than the minimum threshold of 0.5, signifying acceptable convergent validity. The item loadings, composite reliability and AVE are presented in Table 1.

Further, the Fornell-Larcker criterion was utilized to assess discriminant validity which measures the degree to which constructs in a structural model are empirically distinct (Hair et al., 2019). The Fornell-Larcker criterion was used to assess whether the AVE of each construct is more significant than its squared correlation with the remaining constructs (Hair et al., 2019). The square root of the AVE of each construct is indicated with diagonal values in italics with the off-diagonal values representing construct interrelation as can be seen in Table 2. Diagonal values greater than the off-diagonal values shown in Table 2, signify that the model possesses discriminant validity.

Table 2: Fornell-Larcker Criterion

Constructs	Mean	SD	I	II	III	IV	V	VI	VII	VII
CR	4.219	1.369	0.788							
EP	4.711	1.200	0.342	0.862						
II	4.653	1.158	0.419	0.331	0.794					
IC	4.365	1.270	0.410	0.349	0.453	0.817				
IM	4.857	1.194	0.361	0.374	0.656	0.636	0.851			
MBE	4.806	1.356	0.313	0.248	0.116	0.167	0.163	0.808		
SP	5.167	1.012	0.132	0.354	0.343	0.258	0.352	0.164	0.805	
SR	5.167	1.012	0.083	0.229	0.331	0.232	0.220	0.336	0.334	0.759

Before examining the structural model results, correlation analysis was performed to investigate the direction and strength of the relationship between the constructs. The analysis revealed that all the study constructs are related positively though their significance level differ. Only the relationships between inspirational motivation and idealized influence and management-by-exception and individualized consideration were insignificant at the 0.001 and 0.05 significance levels. The correlation coefficients are presented in Table 3.

Table 3: Correlation Coefficients

Constructs	II	IS	IC	CR	MBE	FR	SP	EP
CR	1							
EP	.659**	1						
II	.452**	.633**	1					
IC	.414**	.350**	.420**	1				
IM	.089	.143*	.132*	.288**	1			
MBE	.325**	.235**	.231**	.099	.348**	1		
SP	.347**	.363**	.234**	.137*	.163**	.330**	1	
SR	.342**	.394**	.356**	.381**	.245**	.242**	.355**	1
**. Correlation is significant at the 0.01 level (2-tailed).								
*. Correlation is significant at the 0.05 level (2-tailed).								

Structural model results

After examination of the measurement model in relation to reliability and validity of the study's constructs, the structural model assessment was examined. The structural model results revealed an R^2 of 0.247 and 0.201 for economic performance and social performance respectively. This signifies that 24.7% and 20.1% of the attributable variations in economic and social performance are accounted for by transformational leadership, transactional leadership and firm resilience. To assess the predictive relevance of the model, an additional evaluation of the model fit was carried out using the blindfolding analysis. Predictive relevance assesses the adequacy of a model to predict the observed indicators of every latent construct (Hair et al., 2019). To evaluate the predictive relevance of the model, the Stone-Geisser Q^2 (cross-validated redundancy) was performed using the blindfolding procedure in PLS. Models with Q^2 values greater than 0 are deemed as having predictive relevance (Hair et al., 2019). Lastly, the structural model was estimated using the nonparametric bootstrapping with 5,000 resamples for the examination of the statistical significance in addition to the relevance of the path coefficients (Hair et al., 2019). The hypotheses tests results are presented in Table 3

Table 4: Hypotheses Results

Hs	Relationship	Path Co-efficient	T Statistics	P Values	Decision
NA	TFL -> FR	0.220	3.012	0.002	Supported
NA	TSL -> FR	0.189	2.691	0.007	Supported
NA	FR -> SP	0.223	3.816	0.000	Supported
NA	FR -> EP	0.083	1.205	0.229	Not Supported
H1a	TFL -> SP	0.291	3.340	0.001	Supported
H1b	TFL -> FP	0.251	3.143	0.002	Supported
H2a	TSL -> SP	0.069	0.806	0.421	Not Supported
H2b	TSL -> FP	0.205	2.781	0.006	Supported
H3a	TFL -> FR -> SP	0.049	0.022	2.193	Supported
H3b	TFL -> FR -> EP	0.018	0.017	1.063	Not Supported
H4a	TSL -> FR -> SP	0.042	0.018	2.294	Supported
H4b	TSL -> FR -> EP	0.016	1.070	0.285	Not Supported

Three out of four Not Hypothesized (NH) paths were supported. First, both transformational (β = 0.220, t = 3.012, $p < 0.05$) and transactional leadership (β = 0.189, t = 2.691, $p < 0.05$) had a significant link with firm resilience. Second, firm resilience had a significant link with social performance (β = 0.223, t = 3.816, $p < 0.05$) but an insignificant link with economic performance (β = 0.083, t = 1.205, $p > 0.05$). For the hypothesized paths, the study found support for five hypotheses out of a total of eight hypotheses. First, transformational leadership had a significant link with social performance (β = 0.291, t = 3.340, $p < 0.05$) and economic performance (β = 0.251, t = 3.143, $p < 0.05$), thereby providing support for H1a and H1b. Further, transactional leadership had a positive but insignificant link with social performance (β = 0.069, t = 0.806, $p > 0.05$), but had a positive and significant link with economic performance (β = 0.205, t = 2.781, $p < 0.05$). For the mediation results, firm resilience had a significant mediating effect in the link between transformational leadership and social performance (β = 0.022, t = 2.193, $p < 0.05$); however, firm resilience had an insignificant mediating effect in the link between transformational leadership and economic performance (β = 0.017, t = 1.063, $p > 0.05$). Lastly, firm resilience had a significant mediating effect in the link between transactional leadership and social performance ($\beta = 0.018$, $t = 2.294$, $p < 0.05$); however, firm resilience had an insignificant mediating effect in the link between transactional leadership and economic performance. Annual revenue as a control variable had a significant link with economic performance ($\beta = 0.128$, $t = 2.188$, $p < 0.05$) but an insignificant link with social performance ($\beta = -0.064$, $t = 1.055$, $p > 0.05$).

Discussions

The findings that transformational leadership has a positive and significant relationship with social performance. From the SET perspective, this result suggests that transformational leadership can be leveraged to improve employee welfare, reduce customer complaints, improve customer welfare, and achieve social performance. Similarly, the study's findings highlighted that transformational leadership is linked significantly to economic performance. The finding is similar to prior findings which revealed that transformational leadership is related significantly with sustainable procurement and sustainability performance (Roman, 2017; Burawat, 2019). The positive result demonstrates that a buying firm's transformational leadership can increase organizational innovativeness and improve the organisation's financial performance (Mokhtar et al., 2019a).

Further, the study revealed that transactional leadership has an insignificant link with social performance. Our insignificant finding suggests that the provision of transactional leadership by buying firms does not result in improved social performance. It is probable that transactional leadership indirectly contributes to social performance through some mechanism(s). Transactional leadership utilizes the principle of rewards and punishment towards the accomplishment of goals, thus, it may not be ideal for accomplishing long-term goals such as social sustainability which requires a broader and holistic approach. On the contrary, the study revealed that transactional leadership has a significant link with economic performance. Supplier compliance with requirements can facilitate the acquisition of right quality products to facilitate production and to address the needs of varied stakeholders which can contribute to economic performance (Birasnav et al., 2015). The findings signify that transactional leadership will involve selecting suppliers based on performance metrics such as cost, quality and timeliness which can crucially drive leading firm effort at achieving economic performance.

In addition, to examine the indirect or intervening role of firm resilience in the link between transformational leadership and sustainability outcomes, mediation was tested. The mediating role of firm resilience in the link between transformational leadership and the two sustainability outcomes revealed differing results. Specifically, firm resilience mediated the relationship between transformational leadership and social performance but had an insignificant mediating role in the link between transformational leadership and economic performance. Considering that the direct link between transformational leadership and social performance was insignificant, it is concluded that firm resilience has a full mediation role in the link between transformational leadership and social performance. However, it is concluded that firm resilience has no mediating role in the link between transformational leadership and economic performance as the mediation or specific indirect effect was found to be insignificant.

Similarly, mediation role of firm resilience in the link between transactional leadership and the two sustainability outcomes revealed differing results. Specifically, firm resilience mediated the relationship between transactional leadership and social performance as hypothesized but had an insignificant mediating role in the link between transactional leadership and economic performance. Considering that the direct link between transactional leadership and social performance was significant, it is concluded that firm resilience partially mediates the link between transactional leadership and social performance. However, it is concluded that firm resilience does not mediate between transactional leadership and economic performance. In all, the mediation results signify that the development of firm resilience constitute an important avenue through which firms can channel the contribution of both transformational and transactional leadership in improving social performance.

Theoretical Implications

First, the study enriches the literature on SCL and sustainability performance. Specifically, the study provides valuable insights into the differing contribution of SCL (transformational and transactional) to sustainability performance (social and economic). Such finding is significant as literature demonstrates that the contribution of SCL to sustainability outcomes comprising of social and economic outcomes remains underexplored, and thus necessitates future empirical research to clarify the relationship (Gosling et al., 2014; Blome et al., 2017; Chen et al., 2021).

Seond, the study contributes to literature by empirically validating the qualitative and conceptual insights on SCL and sustainability performance. Also, the study augments the clarity concerning the SCL styles that exert the most significant benefits to specific performance improvements (Chen et al., 2021), such as social and economic performance. The study also contributes to literature by addressing the minimal research that focuses on the simultaneous application of transformational and transactional leadership in SCL research (Mokhtar et al., 2019b).

Further, the study contributes to literature by addressing the lack of clarity on whether the SCL-sustainability performance relationship is direct or indirect and what variables might intervene in these relationships (Sundram et al., 2016) given the mixed and inconsistent findings between SCL and sustainability performance (Foo et al., 2021; Huo et al., 2021). Thus, the study contributes to SCL and sustainability performance literature by theorizing and validating firm resilience as a mechanism through which SCL partly contributes to sustainability performance.

Lastly, the study uses data from a unique empirical setting, specifically Sub-Saharan Africa, to broaden the contextual domain of SCL and sustainability research.

Practical Implications

First, the study provides insights in guiding managers and firms in their quest to achieve sustainability outcomes, particularly in the pharmaceutical and private healthcare sectors. The study thus provides insights that can guide managers in prioritizing, understanding and choosing an appropriate SCL style that can be best leveraged to ensure improvement in social and economic sustainability outcomes in relation to their buyer-supplier exchanges. Such understanding and choice of leadership can ensure innovativeness, creativity, optimal support and compliance from upstream actors in the pursuit of supply chain sustainability goals.

Further, empirical validation of how SCL contributes to sustainability outcomes is imperative and informs firms and managers on the prioritization and development of firm resilience capability. Specifically, the study provides guidance to managers to develop upstream resilience capabilities. Therefore, top managers should provide resources and the needed support for developing resilience. The continuity of pharmaceutical supplies due to upstream resilience can contribute in ensuring that SCL is leveraged to improve sustainability outcomes.

Limitations of the study and suggestion for future research

The study has some limitations which can inform subsequent research. First, data used for the analysis were obtained from pharmaceutical and private healthcare facilities. Many of such pharmaceutical firms are relatively small firms. Thus, subsequent researchers can obtain survey data from different sectors such as manufacturing. Additionally, the survey data was gathered from a developing nation in Sub-Saharan, potentially limiting the ability to generalize the findings to significantly different contexts. Thus, it will be interesting to investigate whether the findings of our study are consistent within significantly different contexts.

Further, a cross-sectional survey approach was utilized, limiting the ability to draw causal inferences from the study. Thus, future studies can investigate the model using a longitudinal approach. Considering that firm resilience mediates the contribution SCL on only social performance, subsequent research can explore other possible intervening mechanisms that can transform SCL into sustainability outcomes in addition to possible contingency factors that can influence the relationship between SCL and sustainability performance.

Finally, future researchers can also explore different forms of SCL and the environmental dimension of sustainability performance to provide holistic insights into how leadership contributes to sustainability performance.

References

Agi, M. A., & Nishant, R. (2017). Understanding influential factors on implementing green supply chain management practices: An interpretive structural modelling analysis. Journal of environmental management, 188, 351-363.

Ali, A., Mahfouz, A., & Arisha, A. (2017). Analysing supply chain resilience: integrating the constructs in a concept mapping framework via a systematic literature review. Supply Chain Management: An International Journal. 22(1), pp. 16-39.

Armstrong, J. S., & Overton, T. S. (1977). Estimating nonresponse bias in mail surveys. Journal of marketing research, 14(3), 396-402.

Asamoah, D., Agyei-Owusu, B., & Ashun, E. (2020). Social network relationship, supply chain resilience and customer-oriented performance of small and medium enterprises in a developing economy. Benchmarking: An International Journal, 27(5), 1793-1813.

Asamoah, D., Nuertey, D., Agyei-Owusu, B. and Akyeh, J. (2021). The effect of supply chain responsiveness on customer development. The International Journal of Logistics Management, 32(4), pp. 1190-1213.

Autry, C. W., & Griffis, S. E. (2008). Supply chain capital: the impact of structural and relational linkages on firm execution and innovation. Journal of business logistics, 29(1), 157-173.

Birasnav, M., Mittal, R., Loughlin, S., 2015. Linking leadership behaviors and information exchange to improve supply chain performance: a conceptual model. Global J. Flex. Syst. Manag. 16 (2), 205–217.

Blome, C., Foerstl, K., and Schleper, M. C. (2017). Antecedents of green supplier championing and greenwashing : An empirical study on leadership and ethical incentives. Journal of Cleaner Production, *152*, 339–350. https://doi.org/10.1016/j.jclepro.2017.03.052

Burawat, P. (2019). The relationships among transformational leadership, sustainable leadership, lean manufacturing and sustainability performance in Thai SMEs manufacturing industry. 36(6) pp. 1014-1036

Chen, L., Li, T., & Zhang, T. (2021). Supply chain leadership and firm performance: A meta-analysis. International Journal of Production Economics, 235, 108082.

Chowdhury, M. M. H., and M. Quaddus. 2017. "Supply chain resilience: conceptualization and scale development using dynamic capability theory." International Journal of Production Economics 188; 185-204.

Cohen, J. (2013), Statistical Power Analysis for the Behavioral Sciences, Routledge, Hillsdale, NJ.

Cohen, W. and Levinthal, D. (1990). Absorptive capacity: a new perspective on learning and innovation. Administrative Science Quarterly. 35(1), pp. 128-152.

Cortez, R. M., Johnston, W. J., & Dastidar, A. G. (2020). The Coronavirus crisis in B2B settings: Crisis uniqueness and managerial implications based on social exchange theory. Industrial

Marketing Management, pp. 125 to 135

Dabhilkar, M., Birkie, S. E., and Kaulio, M. (2016). Supply-side resilience as practice bundles: a critical incident study. International Journal of Operations and Production Management. 36(8), pp. 948-970

Defee, C. C., Stank, T. P. T., and Esper, T. (2010). Performance implications of transformational supply chain leadership and followership, *40*(10), 763–791.

Essuman, D., Boso, N., & Annan, J. (2020). Operational resilience, disruption, and efficiency: Conceptual and empirical analyses. International Journal of Production Economics, 229, 107762.

Fan, Y., and Stevenson, M. (2020). Supplier-initiating risk management behaviour and supply-side resilience : the effects of interpersonal relationships and dependence asymmetry in buyer-supplier relationships, *40*(7), 971–995.

Fiksel, J., Polyviou, M., Croxton, K.L., Pettit, T.J., (2015). From risk to resilience: learning to deal with disruption. MIT Sloan Manag. Rev. 56 (2), 79–86.

Foo, P. Y., Lee, V. H., Ooi, K. B., Tan, G. W. H., & Sohal, A. (2021). Unfolding the impact of leadership and management on sustainability performance: Green and lean practices and guanxi as the dual mediators. Business Strategy and the Environment, 30(8), 4136-4153.

Fontoura, P., and Coelho, A. (2020). The influence of supply chain leadership and followership on organizational performance An empirical study on a Portuguese energy supplier, *15*(3), 333–353.

Gosling, J., Jia, F., Gong, Y., & Brown, S. (2016). The role of supply chain leadership in the learning of sustainable practice: toward an integrated framework. Journal of Cleaner Production, 137, 1458-1469.

Gu, M., Yang, L., and Huo, B. (2021). The impact of information technology usage on supply chain resilience and performance: An ambidexterous view. International Journal of Production Economics, 232, 107956.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. European business review, 31(1), 2-24.

Homans, G., 1958. Social behavior as exchange. *American Journal of Sociology* 63, 597–606.

Hubbard, G. (2009). Measuring organizational performance: beyond the triple bottom line. Business Strategy and the Environment, 18(3), 177-191.

Huo, B., Wang, K., & Zhang, Y. (2021). The impact of leadership on supply chain green strategy alignment and operational performance. Operations Management Research, 14, 152-165.

Lambe, C. J., Wittmann, C. M., and Spekman, R. E. (2001). Social exchange theory and research on business-to-business relational exchange. Journal of Business-to-Business Marketing, 8(3), 1–36.

Lamsam, N. and Charoensukmongkol, P. (2022), “Effect of CEO transformational leadership on organizational ethical culture and firm performance: the moderating effect of competitive intensity”, *Journal of Asia Business Studies*, 17(3), pp. 539-558.

Lorentz, H., Laari, S., Meehan, J., Eßig, M., & Henke, M. (2021). An attention-based view of supply disruption risk management: balancing biased attentional processing for improved resilience in the COVID-19 context. International Journal of Operations & Production Management, 41(13), 152-177.

Liu, W., Wei, W., Choi, T., and Yan, X. (2022). Impacts of leadership on corporate social responsibility management in multi-tier supply chains. *European Journal of Operational Research*, *299*(2), 483–496.

Mokhtar, A. M. R., Genovese, A., Brint, A., and Kumar, N. (2019a). Improving reverse supply chain performance : The role of supply chain leadership and governance mechanisms. *Journal of Cleaner Production*, *216*, 42–55.

Mokhtar, A. M. R., Genovese, M., Genovese, A., Brint, A., and Kumar, N. (2019b). Supply chain leadership : A systematic literature review and a research agenda. *Intern. Journal of Production Economics*, *216*, pp. 255–273.

Mzembe, A.N., Lindgreen, A., Maon, F., Vanhamme, J., (2016). Investigating the drivers of corporate social responsibility in the global tea supply chain: a case study of eastern produce limited in Malawi. Corp. Soc. Responsib. Environ. Manag. 23 (3), 165–178.

Boons, F., & Berends, M. (2001). Stretching the boundary: the possibilities of flexibility as an organizational capability in industrial ecology. Business strategy and the environment, 10(2), 115-124.

Pournader, M., Rotaru, K., Kach, A. P., & Razavi Hajiagha, S. H. (2016). An analytical model for system-wide and tier-specific assessment of resilience to supply chain risks. Supply Chain Management: An International Journal, 21(5), 589-609.

Roman, A. V. (2017). Institutionalizing sustainability : A structural equation model of sustainable procurement in US public agencies. *Journal of Cleaner Production*, *143*, 1048–1059.

Sancha, C., Gimenez, C., & Sierra, V. (2016). Achieving a socially responsible supply chain through assessment and collaboration. Journal of Cleaner Production, 112, 1934-1947.

Silvestre, B. S. (2015a). A hard nut to crack! Implementing supply chain sustainability in an emerging economy. *Journal of Cleaner Production*, *96*, 171–181.

Singh, N.P., 2020. Managing environmental uncertainty for improved firm financial performance: the moderating role of supply chain risk management practices on managerial decision making. International Journal of Logistics Research and Applications 23 (3), 270–290.

Sundram, K., Chandran, V. and Awais, M. (2016), “Supply chain practices and performance: the indirect effects of supply chain integration”, Benchmarking: An International Journal, Vol. 23 No. 6, pp. 1445-1471, doi: 10.1108/BIJ-03-2015-0023.

Terpend, R., and Krause, D. R. (2015). Competition or cooperation? Promoting supplier performance with incentives under varying conditions of dependence. Journal of Supply Chain Management, 51(4), 29-53.

Verghese, A. J., Koufteros, X., Polyviou, M., & Jia, X. (2022). In pursuit of supplier resilience: The explanatory role of customer leadership style. Transportation Research Part E: Logistics and Transportation Review, 159, 102626.

Vivaldini, M., & Pires, S. R. (2016). Sustainable logistical operations: the case of McDonald's biodiesel in Brazil. International Journal of Logistics Systems and Management, 23(1), 125-145.

Wang, M. , Wang, Z. , Fang, X. , and Chen, Y. J. (2019). Impacts of distributional social comparison behavior on corporate social responsibility: Power of the small. Available at SSRN 3350157

Zhu, Q., Liu, J., & Lai, K. H. (2016). Corporate social responsibility practices and performance improvement among Chinese national state-owned enterprises. International Journal of production economics, 171, 417-426.

DECISION SCIENCES INSTITUTE

How Does Transparency Impact Technological Novelty? Evidence From Large Pharmaceutical Firms

Hanu Tyagim University of Minnesota, tyagi035@umn.edu
Manuel Hermosilla, Johns Hopkins University, mh@jhu.edu
Rachna Shah, University of Minnesota, shahx024@umn.edu

ABSTRACT

How does transparency into a firm's new product development initiatives impact technological novelty? We examine the relationship in the pharmaceutical industry where historically, firms hid clinical trial results that showed poor safety or efficacy. We study the impact using ICMJE, an exogenous policy designed to increase transparency. Using a difference-in-differences approach and fine-grained data from 10,000 clinical trials conducted between 2000 and 2014, we find that transparency increases firms' pursuit of less novel, less risky existing technologies. Our results show that transparency negatively impacts technological novelty, and the effect is more pronounced in firms with less diversified portfolios.

KEYWORDS: Technological novelty, Transparency, New product development, Pharmaceutical Industry

1. INTRODUCTION

Managers often prefer less risky options over more risky ones to avoid potential losses, even if the latter are associated with larger expected payoffs (Tversky and Kahneman, 1992; Kahneman and Tversky, 1973). This is particularly true in the new product development (NPD) context, where choice sets are often large, the likelihood of failure is high, and the risk-reward payoffs are uncertain (Kahneman and Lovallo, 1993). In such contexts, decision-makers can become more risk-averse, especially if they are likely to be held accountable for their choice (Kahneman and Lovallo, 1993; Viscusi et al., 1987). This suggests that when managerial decisions are open to public view, the ensuing potential failure would lead managers to choose safer, justifiable options. The specter of potential failure may have a particularly important influence on managerial risk-taking when NPD efforts are transparent to the public. Since novelty often implies a higher level of risk (Dranove et al., 2022), a greater reluctance towards risk-taking could reduce managerial effort toward novel NPD initiatives. To reduce the likelihood of failure, firms may choose to exploit previously used technologies that leverage existing knowledge over exploring novel technologies that provide competitive advantage and superior financial performance but are more prone to failure. A competing hypothesis posits that increased NPD visibility could lead a firm to pursue novel technologies by embracing the competing firm's novel technologies into its own ongoing NPD efforts or using them to justify its own technological choices. Thus, transparency could act as a catalyst for the focal firm's own novel NPD initiatives. We examine this tension by studying whether transparency into NPD activities helps or deters the novelty of NPD projects in technology firms.

In this paper, we investigate the relationship between transparency and novelty in the context of drug development by large pharmaceutical firms. Two primary features make this context an appropriate empirical setting to study our research question. First, the drug development process has been historically characterized by a severe lack of transparency (Miller et al., 2017). Although pharmaceutical firms are required to conduct clinical trials to assess the safety and efficacy of a new drug, they face strong disincentives to publicly disclose clinical trial results that show poor safety or efficacy of the trial drug. The large potential harm to patients implied by such selective reporting creates an inevitable need for a more transparent drug development process. Second, drug novelty is a critical driver of performance for pharmaceutical firms. Novel drugs have the potential to become blockbusters - the primary source of competitive advantage and growth in the industry. However, industry experts bemoan the paucity of novel drugs and project a worrisome outlook for pharmaceutical R&D (Munos and Chin, 2011), primarily driven by constricted pipelines of large pharmaceutical firms (Juliano, 2013). The far-reaching implications of stagnant innovation provide an impetus to examine whether initiatives implemented to benefit key stakeholders have negatively, albeit unintentionally, impacted technological novelty.

Our analysis exploits a large pro-transparency policy shock that affected the pharmaceutical industry. In 2005, the International Committee of Medical Journal Editors (ICMJE) instituted a policy to address biased reporting of clinical trial results by pharmaceutical firms. The policy was designed to create greater transparency by requiring clinical trial investigators to register their clinical trials in a public trials registry as a condition of consideration for publication in academic journals (De Angelis et al., 2004). This quasi-voluntary mandate proved influential as more than 5,000 academic journals (including the most reputed medical journals such as JAMA and The Lancet) abided by the new policy. The sharp increase in clinical trial registration rates following the policy reflects the increased transparency in drug development (Zarin et al., 2005). Although the policy affected every firm in the pharmaceutical industry, the effect varied depending on firms' publication intensities before the ICMJE policy enactment – firms with high publication intensities may have been more exposed to the policy compared to firms with low publication intensities.

To investigate our question, we quantify the technological novelty of drugs tested in approximately ten thousand clinical trials sponsored by large pharmaceutical firms between 2000 and 2014. We construct this novelty metric based on fine-grained information about the pharmacological approach (“technology”) used by each drug. Specifically, we measure drug novelty at the trial level, as the number of previously tested drugs that use the same technology as that used by the drug tested in the trial. We label the resulting variable “prior deployments” where lower values indicate higher levels of technological novelty. At the extreme, zero prior deployments indicate a technology that has never been used before and hence denotes maximum novelty.

Our main econometric analysis uses a differences-in-differences specification that exploits different levels of academic publishing intensity among firms before and after the ICMJE policy enactment. Our data also indicate considerable variation in publishing intensities among firms, where some firms publish less than one academic article for each drug entering clinical trials while others publish almost forty. Our results show that the additional transparency infused by the ICMJE policy negatively impacts the technological novelty of drugs. Specifically, our estimates suggest that one standard deviation higher policy exposure is associated with a 13.9% increase in the average number of prior deployments, implying a shift towards less novel, less risky projects. Interpreting this estimate in light of the average rate of technology deployment accumulation in our data suggests that higher policy exposure results in firms falling behind the technological frontier by approximately two years.

A potential mechanism behind the shift towards less novel projects can be attributed to negative informational spillovers. In the context of drug development, such spillovers manifest when the failure of one technology leads to negative inferences about the viability of the failed technology in other drugs, or in other related technologies (Krieger, 2021; Magazzini et al., 2012). Since transparency makes failures more difficult to hide from public purview, firms may turn to less novel technologies, which are inherently less risky, to reduce the prospective burden of such effects. Further, we posit that less technologically diversified firms may be more vulnerable to negative informational spillovers compared to more diversified ones because each failure impacts a less technologically diverse firm more profoundly than its more diversified counterparts. We examine the moderating role of technological diversification within each firm's portfolio of drugs in development and find that transparency's effect on novelty is indeed driven by firms whose technology portfolios are relatively less diversified.

We also investigate how transparency affects the balance between exploitative and exploratory R&D activities using exploitation-exploration and ambidexterity lenses. While exploitation-exploration implies a trade-off between these two activities (i.e. more of one, implies less of the other), ambidexterity is based on the idea that firms pursue both exploitative and exploratory activities simultaneously to balance the inherently different levels of risks associated with each of them. To examine these perspectives, we classify technologies into previously used technologies that are low-risk (prior deployments > 0) and purely novel technologies that are high-risk in nature (prior deployments = 0). In addition, we identify "near-novel" technologies, i.e., technologies that have been used less than an average technology (prior deployments < average prior deployments), which can be considered medium-risk. Through three separate analyses, we find that although firms with greater exposure to transparency increase exploitation, they do not reduce truly novel pursuits. Instead, they curtail near-novel exploration. Together, these results add nuance to the trade-off-based exploitation-exploration view and the simultaneity-based ambidexterity view.

Our results have important implications for theory, practice, and regulation. Our primary theoretical contribution lies in demonstrating the negative impact of transparency on the technological novelty of the firm, and in empirically illustrating the underlying mechanism - circumventing negative informational spillovers - through which the impact manifests itself (Card et al., 2012; Beer et al., 2021). Our study also highlights firm-level implications of transparency. In doing so, it departs from the vast majority of existing literature which has focused on the individual customer, workers, physicians, etc. Our results add greater granularity to our understanding of the innovation ambidexterity literature (He and Wong, 2004). When evaluated through the exploitation-exploration lens, our results show that the increase in prior deployments is driven by a higher exploitation of previously used technologies but not a lower exploration of new technologies. This suggests that the trade-off-based exploitation-exploration lens might not be adequate to interpret our results. The additional analysis shows that instead of reducing the use of purely novel technologies, firms reduce the use of near-novel technologies. While the results imply the simultaneous pursuit of exploration and exploitation and seem more aligned with ambidexterity in the NPD context, they also indicate a need for greater nuance when invoking ambidexterity. Here, our contribution lies in making a distinction between novel and near-novel activities and supplementing the traditional exploitation-exploration literature. Distinguishing these two categories will allow future researchers to get better insights into managerial decision-making around risk.

Our results related to diversification suggest a potential solution for managers to counteract the negative impact of transparency on technological novelty. Managers could diversify their firm's technology portfolios to reduce vulnerability from transparency shocks. More broadly, our results add to managerial understanding by providing plausible explanations for key trends recently observed in the pharmaceutical industry. Perhaps, the growing concerns around the lack of novel drugs, outsourcing of early-stage R&D, and growth of large companies by acquiring smaller new firms can each be partially explained by increased transparency in the pharmaceutical industry. It is likely that firms have adopted each of these approaches to reduce their exposure to risk.

Finally, for policymakers, we highlight the inherent trade-off between transparency and the novelty of innovation. While increased transparency is a boon in and of itself, it is possible that it may have contributed to the "incremental innovation" problem where pharmaceutical firms prioritize small improvements in their existing portfolio over pursuing novel breakthrough innovations. Our results suggest that policymakers may need to consider financial mechanisms to incentivize firms to undertake risky R&D projects (Krieger et al., 2022; Lo and Thakor, 2022) to ameliorate the downside risks associated with increased transparency.

The rest of the paper is structured as follows. First, we provide a brief review of the relevant literature streams, the contextual details around drug development, and the ICMJE policy in the pharmaceutical NPD context. Next, we describe the data, operationalization of our key variables, and research method. We then describe our main results and provide evidence of the mechanism to corroborate our main results. We conclude by discussing the contributions of our study to management theory and practice.

2. LITERATURE REVIEW

To examine whether transparency into NPD activities positively or negatively affects the technological novelty of a firm, we review literature related to transparency and innovation. Within the innovation stream, we pay special attention to papers that focused on the NPD context. We use literature related to risk as the overarching theoretical grounding for our research question. We provide a summary of our review below.

2.1. Transparency

The literature on transparency is relatively new but has grown fast and has expanded to diverse service contexts such as healthcare, restaurant, and the government sector. Although there is no universally accepted definition of transparency, the concept encapsulates the notion of “visibility” and “ability to observe or look into” how a task or an activity is performed, how much effort is expended, what are the task and process outcomes, etc. We find that the activity being observed (e.g., tasks, effort, outcomes, etc.), the mode of observing the activity (e.g., self or mandated disclosure, surveillance, monitoring, etc.), and the associated individual, group, and organization level outcomes (e.g., productivity, learning, satisfaction, compliance) are well-developed themes within this literature [Footnote: For a comprehensive review of this multidisciplinary literature, see Bernstein (2007)].

We summarize a representative sample of this growing literature in Table 1 and highlight several interesting characteristics that we gleaned in our review. We find that the unit of analysis in a vast majority of studies is an individual, such as customers (Buell and Kalkanci, 2021; Mohan et al., 2019), an employee (Card et al., 2012; Beer et al., 2021; Cullen and Perez-Truglia, 2022), or a physician (Chao and Larkin, 2022; Dranove et al., 2003). We also see that the majority of studies consider transparency initiatives that are voluntary in nature such as the promotion of social responsibility initiatives (Kraft et al., 2018; Buell and Kalkanci, 2021) and firm authenticity (Mejia et al., 2019). In general, we find that increased transparency leads to positive, desirable outcomes, such as higher productivity (Mas and Moretti, 2009), higher service quality (Buell and Norton, 2011; Buell et al., 2017), and higher customer purchase intentions (Kraft et al., 2018; Mohan et al., 2019; Buell and Kalkanci, 2021).

Our study differs from the existing papers in several important ways. To the best of our knowledge, it is the first empirical study that examines how transparency affects the high risk-reward payoff typically associated with novel technologies [Footnote: Brown and Martinsson (2019) studies the impact of transparency on the amount of innovation at a national level. However, we are interested in understanding the firm-level implications of increased transparency on the novelty of innovation]. In contrast to the majority of research that examines transparency in a service setting with an individual as the unit of analysis, our study is conducted at the firm level in the NPD context. Additionally, we measure transparency with an exogenous policy shock that is quasi-voluntary in nature. Specifically, our measure is neither a voluntary initiative nor a regulatorily required mandate (see Section 3.3 for more details). Thus, we move exogenous policy shocks from a categorical binary indicator to a more calibrated variable. Lastly, in examining the potential undesirable effects of transparency, our study differs from others that evaluate the desirable effects of voluntary and involuntary measures (Mas and Moretti, 2009; Chao and Larkin, 2022; Lee et al., 2021). Thus, our study contributes to the transparency literature by examining the potential undesirable impact of a quasi-voluntary measure of transparency on technological novelty in the NPD context at the firm level.

2.2. Innovation

In contrast to transparency, innovation is a well-established and well-studied concept. According to the Merriam-Webster dictionary, innovation is defined as newness (e.g. a new idea, method, or device) and novelty (e.g., the introduction of something new). It has been studied in numerous contexts such as routines and practices used to conduct work, production methods and technologies, products and processes, and even organizational design. A comprehensive review of these distinct literature streams is beyond the scope of this paper. Instead, we highlight the most salient details related to innovation in the NPD context that are pertinent to our study.

In the NPD innovation literature, the notion of novelty and risk are interrelated, and the level of risk is thought to vary with the “extent of the novelty” of innovation. To better capture the varying levels of risk, researchers have made a distinction between radical and incremental innovation (Peng et al., 2008; Chao and Kavadias, 2008; Eggers and Kaul, 2018). Radical innovation involves a clear departure from established norms and is generally disruptive in nature. It requires a firm to develop new competencies which may result in novel products and breakthrough technologies. Researchers often refer to radical innovation as exploration and have empirically shown that exploratory activities provide firms with a first-mover advantage and create entry barriers for competing firms, which enable firms to appropriate higher rents and achieve superior market and financial performance (Chandy and Tellis, 2000; Kirchhoff, 1992; Schumpeter, 2013). Incremental innovation, on the other hand, involves an improvement over the current state rather than a disruptive departure from the prevailing norm (March, 1991; Cheng and Van de Ven, 1996). Incremental innovation allows a firm to leverage its existing knowledge and technologies and build on existing firm-specific resources. Frequently, incremental innovation and exploitation are used interchangeably. From a firm’s perspective, exploitative activities are desirable because leveraging existing competencies and technologies reduces the likelihood of failure, and building on available firm-specific resources results in more cost-efficient product development (Benner and Tushman, 2003; Rothaermel and Deeds, 2004; Chandrasekaran et al., 2015).

Although radical innovation and exploratory activities are considerably more important for sustained competitive advantage, and long-term firm growth and survival, they are also associated with higher levels of risk because new products and technologies have a higher likelihood of failure. Therefore, it is essential for firms to balance high-risk, high-reward exploratory activities with low-risk, low-reward exploitative ones. This is particularly true in a firm’s new product development efforts where firms need to successfully innovate and at the same time ensure that exploiting existing knowledge does not crowd out novel, exploratory initiatives (Benner and Tushman, 2003; Gibson and Birkinshaw, 2004; Uotila et al., 2009). Interestingly, most successful firms undertake both exploratory and exploitative activities, an approach, popularized as ambidexterity (March, 1991; Swift, 2016; Mudambi and Swift, 2014). Given the importance of developing new products to a firm’s future success and the temptation of using established technologies for efficient use of resources, it is critical to understand how external initiatives that enforce increased visibility into a firm’s new product development efforts impact its innovation efforts (Benner and Tushman, 2003).

3. BACKGROUND

We examine the impact of transparency on technological novelty in the pharmaceutical industry. In this section, we briefly describe the role of clinical trials in the drug development process and how transparency manifests in the process. We then describe the ICMJE policy, its main precepts, and how it sought to increase transparency in the drug development process.

3.1. New product development in the pharmaceutical industry

In the United States, the Federal Drug Administration (FDA) regulates drug development and approval and requires pharmaceutical firms (also called “sponsors”) to follow a highly structured stage-gate process (Cooper, 1990). The drug development process begins when a new molecular entity (NMEs, hereafter “drugs”) targeted to a specific disease condition is identified. Subsequently, the drug goes through a series of rigorous tests in animals for toxicity, dosage levels, and absorption during the pre-clinical trial stage. Based on the results, a subset of these drug candidates advances to clinical trials, and a relatively small proportion of drug candidates entering the clinical trials get approved by the FDA. Only drugs that receive FDA approval can be marketed in the US.

Clinical trials are interventional studies where the drug candidate’s pharmacological effects are evaluated on humans. Typically, drugs undergo at least three sequential phases of a clinical trial, referred to as Phase 1, Phase 2, and Phase 3. The phases vary in duration and the evidentiary requirements become increasingly stringent with each passing phase. In comparison to “early-stage” Phase 1 and Phase 2 trials, the “late-stage” Phase 3 trials are usually the final, longest (several months to years), and largest (several hundred to thousands of patients) steps of the process. Given the stage-gate structure (Cooper, 1990), drugs entering the late stage, Phase 3 clinical trials have a higher likelihood of drug approval compared to drugs in early-stage trials.

An important feature of a clinical trial is the underlying “technology” of the tested drug. The technology represents the pharmacological approach that the drug deploys to deliver its therapeutic benefit. It entails fine-grained information such as the biological entity within the human body targeted by the drug, the mechanism used by the drug to target it, etc. It is important to note that each drug could have multiple technologies associated with it, and a single technology could be the basis of multiple drugs. A drug based on a never-before-used technology is associated with higher levels of risk and needs to provide more stringent safety evidence at each clinical trial stage. Several researchers have noted the relationship between technological novelty and risk in the context of drug development (Hara, 2003; Munos and Chin, 2011; Krieger et al., 2022; Dranove et al., 2022). Every successive use of the new technology provides additional information, which incrementally reduces the level of risk associated with that technology. The repeated usage of technology would, therefore, indicate lower novelty.

In general, developing a new drug is a long and expensive process with uncertain outcomes as it may cost a pharmaceutical firm more than 2 billion USD and may take almost 9 years to receive FDA approval [Footnote: https://www.cbo.gov/publication/57025]. Moreover, only about 7% of drugs entering clinical trials receive FDA approval (Dowden and Munro, 2019). Such low odds of success, especially for drugs that involve a novel technology, have made pharmaceutical firms extremely sensitive to disclosing any negative outcomes during clinical trials.

3.2. Transparency in clinical trials

Pharmaceutical firms are not required to publish data related to clinical trial processes and outcomes as it is considered proprietary knowledge under federal and state-level definitions of trade secrets (Kesselheim and Mello, 2007). Trial sponsoring firms determine how, when, and how much information about the clinical trials is to be released in the public domain. One of the primary means for firms to disclose information about the clinical trial and its outcomes is by publishing them in an academic journal (Gans et al., 2017). Publications are an important means for pharmaceutical firms to fend off competitive pressure (Polidoro and Theeke, 2012), enhance productivity in drug discovery (Henderson and Cockburn, 1994), and improve commercial prospects of marketed drugs (Azoulay, 2002). It also allows firms to retain their academically inclined scientists (Zucker and Darby, 1997). Several researchers note that scientific personnel, particularly in the pharmaceutical industry, leave when they are not allowed to publish or not publish to the extent they aspire (Schiermeier, 2010; Stern, 2004). Thus, the ability to publish is a sought-after advantage for the sponsoring firm and the scientists employed by the firms.

Trial sponsoring firms are more likely to publish clinical trial results when the information related to clinical trials is positive and less likely to publish when the trial results unveil safety issues or lackluster efficacy. Poor clinical trial results negatively impact the sponsoring firm's valuation (Girotra et al., 2007) and could raise questions about the drug's quality which could jeopardize its approval and lower consumer demand if approved. Several recent cases illustrate these points: GlaxoSmithKline in the case of its depression drug Paxil and the anti-diabetic drug, Avandia, and Merck in the case of the painkiller, Vioxx. In each of the cases, the sponsoring firm was accused of withholding information pertaining to drugs' safety profiles to avoid negative repercussions.

Negative clinical trial results may also cast aspersions on other products in the firm's portfolio (Krieger, 2021; Magazzini et al., 2012). This negative spillover is particularly evident for drugs that rely on the specific technology used in the test drug, or other related technologies. A case in point is the SMARCA4 ligand, targeted for lung cancer, and the aftermath of the publication of a study debunking its effectiveness (Vangamudi et al., 2015). Commenting on the repercussions, a renowned industry observer wrote "*A number of programs (in academia and industry both) came to a juddering, dust-spewing halt once people saw the data*" (Lowe, 2015).

Such negative spillover effects create strong incentives for pharmaceutical firms against publishing negative clinical trial results. This selective reporting where negative trial results are hidden from public purview is commonly referred to as the "file-drawer" problem in the pharmaceutical industry (Rosenthal, 1979). The biased reporting not only reduces transparency but also raises concerns about endangering patients' safety due to distorted clinical evidence available during treatment (Greenberg, 1999). One of the most egregious and consequential examples of such opaqueness is Paxil, a drug approved by the FDA for treating depression, and subsequently withdrawn from the market because it increased the risks of suicide among teenagers. An internal memo leaked to the public revealed that GlaxoSmithKline, the sponsoring firm, had discovered the risk during clinical trials but hid the information (Kondro, 2004), which led legal authorities in the US and UK to launch investigations based on "repeated and persistent fraud" (Sataline, 2005) and prompted the industry to take pro-transparency actions.

3.3. The 2004 ICMJE transparency mandate

To curb selective reporting of trial results, the International Committee of Medical Journal Editors (ICMJE) launched a policy aimed at increasing clinical trial transparency (Meier, 2004a,b). ICMJE is a consortium of editors of the twelve most influential medical journals worldwide. The policy leveraged academic publishing by scientists employed in pharmaceutical firms and made timely registration of clinical trials a pre-condition to reviewing and publishing research based on trial results. Trial registration required a sponsoring firm to submit fundamental trial information (e.g., study design, protocol, outcome measures, etc.) to a publicly accessible clinical trial database (or a trial registry) before recruiting patients for the trial. The rationale was that trial registration would discourage cherry-picking of results reaching the public domain, and provide healthcare providers and patients with a more balanced picture of drugs' risks and benefits (Brownlee, 2004).

The policy was implemented expeditiously. ICMJE announced the policy in September 2004 and became effective in July 2005 (De Angelis et al., 2004). Trials starting after July 1st, 2005, needed to be registered before the first patient was enrolled; and trials already underway were required to retrospectively register no later than September 13th, 2005. Although pharmaceutical firms reacted strongly citing concerns related to revealing proprietary information (Sataline, 2005; Brownlee, 2004), over 5000 other journals agreed to abide by ICMJE guidelines. The policy had an immediate and persistent impact on trial registration rates. We illustrate the number of clinical trials registered on clinicaltrials.gov [Footnote: Although the policy did not mandate the use of a particular trial registry, clinicaltrials.gov (CTGov) was the only registry that satisfied the minimum features required by the editors] in Figure 1. We find that the number of registrations grew sharply following policy implementation, from an average of approximately 60 trials per quarter in the pre-policy period (Q1, 2000 - Q2, 2005) to 1,484 in Q3, 2005. After the initial rapid increase, registrations stabilized at a rate of about 500 trials per quarter in the post-policy period (Q4, 2005 - Q4, 2014). Zarin et al. (2005) estimate that trial registrations increased by 73% increase between May and October of 2005. Although several other similar policies have been subsequently enacted [Footnote: For the full list of these policies, see https://clinicaltrials.gov/ct2/about-site/history], the ICMJE policy is considered the most important and the most consequential effort towards pro-transparency (Zarin et al., 2017). In our empirical analysis, we use ICMJE policy as the exogenous shock that enhanced industry-wide transparency levels by requiring that all clinical trials be registered to be considered for academic publication.

4. DATA

To examine the impact of transparency on technological novelty, we use the drug development data compiled by Clarivate Analytics' Cortellis - a pharmaceutical intelligence database that tracks drug development activities worldwide. Cortellis collects information on new clinical trials from a variety of sources around the world. To ameliorate concerns related to missing data and late filings, over 500 Cortellis experts expend significant effort to ensure data integrity. For instance, they backfill late trial disclosures to reduce missing data (Krieger et al., 2022). It is plausible that the ICMJE policy led firms to disclose a sizable number of clinical trials after the policy was implemented. Cortellis includes all such late disclosures at the time the trials are announced and also reports the original trial initiation date. Moreover, the time between policy implementation in 2005 and our data collection in 2021 is sufficiently long for Cortellis to backfill any late disclosures. We empirically assess the extent to which the policy may have impacted Cortellis' observability to address concerns around the data. In Figure 1, we plot the time series of trial initiations in the Cortellis dataset (dashed line) against the time series of trial registrations on CTgov (solid line). The underlying hypothesis that the ICMJE policy may have affected Cortellis' ability to collect data would be reflected by a discontinuity around 2005 in the dashed line. Contrary to this hypothesis, the series of Cortellis trial initiations display continuous growth throughout the sample, without any visible discontinuities around the policy implementation period [Footnote: We formally test for discontinuity using the structural break test. One of the methodologies for the structural break test is to follow Bai and Perron (2003). This technique allows the time series to endogenously learn where the breakpoints could be instead of researchers choosing them exogenously. While we find a structural break at the policy implementation date in the CTgov trial registration data, we do not find any structural break around the same date in the Cortellis trial initiation data. This evidence suggests that the policy implementation did not result in missing data].

Over time, Cortellis has gained broad acceptance as the industry's most comprehensive and up-to-date source of clinical trial data and has been recently used in a number of recent works (Dranove et al., 2022; Hermosilla, 2021, 2022; Krieger, 2021; Krieger et al., 2022; Subramanian et al., 2020; Khmelnitskaya, 2021). We construct a database where each observation corresponds to a clinical trial and extract a comprehensive set of variables for each clinical trial, including the name of the tested drug and the sponsoring firm(s), the trial's phase, initiation date, the disease indication for which the drug is being tested, therapeutic area, etc. We also collect information on the drug's underlying technology and use it to construct our measure of technological novelty.

In this study, we focus on interventional trials for novel human therapeutic drugs [Footnote: The sample excludes trials for generics, and biosimilars, as well as those for imaging technologies, diagnostic tests, and medical devices] initiated between 2000 and 2014. Since the policy was implemented on September 13th, 2005, we define the period from January 1st, 2000 to September 13th, 2005 as the "pre-period". We used 2000 as the start year because it provides a sufficiently wide pre-policy time window. We define the "post-period" as the period from September 14th, 2005 to December 31st, 2014. The study duration gives us a relatively long post-policy period of about 9 years, which is desirable from an empirical standpoint given that a drug, on average, takes almost 9 years from the first clinical trial to commercialization [Footnote: https://www.cbo.gov/publication/57025]. Moreover, we focus on pharmaceutical firms that are members of the Pharmaceutical Research and Manufacturers of America (PhRMA), i.e., the consortium of the largest pharmaceutical manufacturers operating in the United States. Focusing on PhRMA manufacturers allows us to exclude small pharmaceutical firms with narrow technological portfolios that may have little scope to decide which technology to develop. For small pharmaceutical firms, transparency may do little by way of changing their technological choice sets in NPD. By contrast, large firms with broad portfolios of developing technologies are known to actively manage their pipelines based on updates to the economic environment (Chan et al., 2007; Ding and Eliashberg, 2002; Girotra et al., 2007) [Footnote: Of the firms that were active PhRMA members during 2021, we exclude 10 firms given data limitations for the following reasons: (i) the firm was founded after the policy change (1 firm), (ii) no trials were reported in the Cortellis data (2 firms), and (iii) no associated academic publication data (7 firms). For the included firms, we harmonized reported firm naming to include all relevant records. For example, trials linked to Sanofi in the constructed dataset include trials that Cortellis links to the sponsors "Sanofi Canada," "Sanofi GmbH," "Sanofi Inc.," among several others. We also link trials sponsored by subsidiaries and joint ventures. For example, trials sponsored by the Sanofi and Bristol-Myers Squibb (BMS) joint venture are assigned to both Sanofi and BMS].

Our final sample consists of 9,723 clinical trials. Figure 2 shows the distribution of trials across therapeutic areas and phases. Oncology is the single largest therapeutic area, accounting for 18.2% of the trials. Neurology and Immunology come in second and third, accounting for 11.5% and 8.6% of the trials respectively. Phase 1, phase 2, and phase 3 trials account for 29.0%, 43.7%, and 27.3% of the trials respectively.

To capture the impact of ICMJE policy, we leverage the academic publication data from Altmetric, a data science company that tracks scholarly publications and their impact. Altmetric database stores data at a journal article level and provides bibliographic information such as the article title, authors, date of publication, journal title, etc. Importantly, Altmetric also provides information on the funding sources for each article. We use this information to link the firms in our sample to the academic research they funded.

5. METHODS

5.1. Research Design

We leverage the exogenous enactment of the ICMJE policy to investigate how enhanced transparency affects firms' willingness to invest in risky NPD projects. Since the policy affected every firm in the pharmaceutical industry, albeit variably, there are no treated or control groups making the canonical difference-in-differences estimation approach infeasible in this context. Therefore, we adopt a difference-in-differences estimation method in which, instead of relying on separate treatment and control groups, the causal effect of a treatment is estimated by exploiting the heterogeneity in the intensity of treatment among units. In our context, the impact of increased transparency is likely to vary significantly across firms. The ICMJE policy is likely to impact firms that, before the policy enactment, published with high intensity a lot more compared to firms that published with relatively lower intensity. Our econometric framework leverages this variation as the heterogeneous intensity of treatment among firms. We incorporate this variation through a "policy exposure" variable, which tracks the firm-level pre-policy intensity to publishing relative to the number of drug candidates brought to clinical trials. A higher policy exposure would mean a higher intensity of treatment. We combine this cross-sectional policy exposure variation with the timing of the policy event into a difference-in-differences model with varying treatment intensity. A similar approach has been used by other researchers (Acemoglu et al., 2004; Duggan and Morton, 2010; Parker et al., 2016; Hermosilla et al., 2018).

5.2. Variables

5.2.1. Policy Exposure

The ICMJE mandate sought to increase clinical trial transparency by precluding the publication of articles that report the results from unregistered trials. Therefore, firms that are more involved in academic publishing should have been more affected by the policy. Based on this observation, we construct a policy exposure variable that quantifies the importance of academic publishing for each firm. For each firm i, we define a policy exposure variable as:

$$Exposure_i = Standard\ Normal\left(\frac{Publications_i}{Drugs_i}\right) \tag{1}$$

The numerator (Publications) in Equation 1 corresponds to the number of articles published in medical or life-science journals that reported funding from firm i. We compute this quantity using the Altmetric database, by counting all academic articles funded by the firm [Footnote: In case an article has multiple funding sources, we count the article under all the associated funding organizations.]. In turn, we compute the denominator (Drugs) using Cortellis data, by counting the number of new drugs that the firm introduces to clinical trial development. To avoid contamination with the policy's effects, both variables (numerator and denominator) are measured using data from before the policy's creation, i.e., from January 1st, 2000 to September 13th, 2005. We then standard-normalize the publications to drugs ratio to facilitate the interpretation of our key estimates. As a result of this transformation, the exposure variable has a mean of zero and a standard deviation of one.

Note that the variable computed in Equation 1 assigns larger policy exposure to firms that published more articles per drug. For firms with minimal publishing, the policy's enactment may have done little by way of changing incentives. By contrast, for firms with high publication intensities, failure to comply with the policy would result in considerable disruptions to their intellectual endeavors. As such, we expect the firms with high publication intensities to have more prominently adjusted their innovation decisions in response to the new policy.

5.2.2. Technological Novelty

To construct our dependent variable, we use the descriptions of the technology used by each drug in the Cortellis database. These descriptions systematically codify the pharmacological approach employed by the drug tested in each clinical trial. Each technological description contains two pieces of information: (i) the targeted biological entity and (ii) the mechanism used to target it. For example, the technology described as "5-HT2 receptor antagonist" acts by antagonizing (i.e., blocking the response of) 5-HT2 receptor proteins. Similarly, "5α-reductase inhibitor" describes a drug acting by inhibiting the function of 5α-reductase enzymes. This approach to pharmacological classification spans a broad technological space - almost 8,000 unique technologies in the full Cortellis dataset going as far back as the mid-1970s (1,307 covered by our data duration) [Footnote: In the Cortellis dataset, the technological descriptions used by our analysis are called "target-based actions." For ease of exposition, we simply refer to these as "technologies".]

We label our measure of technological novelty "prior deployments." For a trial j testing drug d, we define prior deployments as the number of other drugs previously tested in clinical trials that rely on the same technology as drug d. [Footnote: While most drugs in our sample (70%) are associated with a single technology, some are associated with more than one (20% with 2, 5% with 3, 5% with >3), with an average of about 1.5 technologies per drug. For the 30% of drugs associated with more than one technology, we average prior deployments across all listed technologies.] To illustrate the construction of the variable, consider the following two examples. When Astra Zeneca initiated a Phase 3 trial for its drug Quetiapine, the drug's associated technology (5-HT2 receptor antagonist) had been previously used by 48 other drugs tested in clinical trials (as recorded by the full Cortellis dataset). Accordingly, we associate this trial with prior deployments=48. In turn, when GlaxoSmithKline started a Phase 2 trial for its drug Izonsteride in 2001, the associated technology (5α-reductase inhibitor) had been previously used by 11 other drugs tested in clinical trials (prior deployments=11). Because the latter trial tested a drug with fewer prior deployments, we can deem it a more risky and more novel development activity than the former. Each deployment would produce some risk-reducing information about the technology which makes incremental use of the technology less novel. Therefore, we associate smaller values of prior deployments with higher levels of risk and novelty.

An implicit assumption behind our measure of technological novelty - prior deployments – is that all deployments produce the same amount of risk-reducing information. However, some deployments could offer better information and thereby, may be considered less risky and less novel than others. For example, consider prior deployments of a technology that resulted in the successful commercialization of the drug. Successive deployment of a technology that was "validated" through passing phase-wise clinical trials would result in greater risk reduction than others and could be deemed as less novel than others. We incorporate this heterogeneity into our analysis in Section 6.4.3.

5.3. Econometric model

We estimate the following difference-in-differences specification with prior deployments as the dependent variable and Exposure as the continuous treatment variable:

$$\mathbb{E}[Prior\ Deployments_j] = f(\beta * Exposure_{i(j)} \times Post_j) + \emptyset * X_j + \lambda_{i(j)} + \tau_{year(j)} + \mu_{phase(j)}) \quad (2)$$

where j indexes trials and Exposure is as defined in Equation 1. We use the functional subindex $i(j)$ for firms (each trial maps to a single firm). The variable Post is an indicator for trials initiated after the policy's implementation, i.e., $Post = 1[Trial\ i\ initiated\ after\ September\ 13th\ 2005]$. Given that the dependent variable - prior deployments - is a count, we estimate the equation with a Poisson quasi-maximum likelihood regression which is an apt choice when the dependent variable is an over-dispersed count variable. We cluster standard errors at the firm level.

We include a series of control variables in the equation. First, the vector X contains therapeutic area dummy variables as well as an indicator for whether the tested candidate is a biological or a chemically synthesized agent. In addition, the specification includes firm-, year-, and phase-specific fixed effects (respectively, λ, τ, μ). Note that the addition of firm- and year-fixed effects would make the inclusion of the Exposure and Post variables redundant.

The parameter of interest in Equation 2 is β (associated with Exposure×Post), which represents the difference-in-differences effect. A $\hat{\beta} > 0$ would indicate a disproportionate percentage increase of prior deployments, thereby, denoting less novelty after the ICMJE policy among firms with higher policy exposure.

6. RESULTS

6.1. Descriptive statistics

In this section, we provide the main empirical results, check the robustness of our results, and test the underlying mechanism. We begin by providing descriptive statistics and model-free evidence below. Our unit of analysis is a clinical trial. Our data sample consists of 9,723 clinical trials and 2,888 drugs during our total study window. Table 2 provides firm-level descriptive statistics for the main variables included in our study. For clarity, we also show the period over which we compute each variable.

6.1.1. Exposure

Over the pre-policy period, firms in our sample were associated with a total of 7,571 academic publications and introduced 808 drugs. The raw and normalized exposure values in rows 4 and 5 of Table 2, respectively, show significant heterogeneity across firms. Positive (negative) values of the normalized Exposure metric correspond to firms with above-average (below-average) publication propensities per drug entering clinical trials. This variation could be explained by a variety of organizational practices. For example, Arora et al. (2018) proposes a conceptual framework positing that the decision of publishing hinges on the firm-specific evaluation of the costs and benefits of publishing. On the one hand, publishing propensity could be negatively impacted by the reduction in the firm's investments in internal scientific research, the risk of informational spillovers to rival firms, or the inability of the firm to patent the knowledge later. On the other hand, publishing could offer reputational advantages (Polidoro and Theeke, 2012), strengthen the firm's competitive position (Azoulay, 2002), help firms collaborate with biotechnology firms, and strengthen ties with the scientific community (Lerner and Malmendier, 2010). Furthermore, firms that publish more use publishing as a perk to scientists. Firms often have incentive schemes to affect the publication output through their scientists (Bhaskarabhatla and Hegde, 2014). Pharmaceutical firms, in particular, use varying degrees of publication-friendly policies to recruit and retain scientists (e.g., Zucker and Darby, 1997), thereby, resulting in the varying propensity to publish.

6.1.2. Technological novelty

We measure technological novelty by the number of prior deployments of technology. Prior deployments has an average of 16.3 (Table 2, row 6) implying that in our sample an average technology is used approximately 16 times but varies across firms. We examine its pattern of variation over time and present the results in Figure 3. First, we plot prior deployments by trial stages in Figure 3A. We note that the average prior deployment for early- and late-stage trials shows a positive trend, suggesting that, on average, drugs become less novel over time across both stages. We also find that the prior deployment of late-stage phase 3 trials (solid line) is significantly larger compared to early-stage Phase 1 and Phase 2 trials (dashed line). This suggests that drugs entering late-stage trials are less novel compared to drugs entering early-stage trials, perhaps because novel drugs do not survive through the early-stage Phase 1 and 2 trials and are less likely to proceed to more advanced, late-stage Phase 3 trials. We account for the differences over time and across clinical trial phases by including year- and phase-fixed effects in our econometric model. Next, we plot the proportion of trials with purely novel technologies, i.e., technologies that have not been previously deployed (prior deployments = 0) by phase over time (Figure 3B). Interestingly, we find that such trials are observed throughout the sample and across phases, at a fairly stable rate of approximately 8%.

6.1.3. Prior Deployments as technological lag

We present Figure 3C to estimate the rate at which prior deployments accumulate over time. Each semi-transparent curve represents a technology. We plot the natural logarithm of the number of deployments that the technology accumulated (y-axis) against the number of years after the first deployment (x-axis). The dark curve represents the "technological novelty frontier" - the average rate of deployment accumulation of technologies in our data. A higher deployment accumulation rate indicates that the rate at which technologies are being reused is higher resulting in less novel technologies in trials. Given the log transformation, the linear pattern indicates that deployments accumulate at a roughly constant rate of 6.3% of additional deployments every year. In other words, it takes an average technology approximately 0.16 years (= 1/6.3) to accumulate an additional percentage of deployments. If the transparency shock increases this rate (6.3%), the average technology would take lesser time to accumulate the additional percentage of deployments. This difference in time could be interpreted as the "technological lag" with respect to the novelty frontier. Therefore, we interpret a percentage increase in the number of prior deployments as 0.16 years of technological lag. This interpretation helps us better understand the strategic implications of our empirical estimates.

6.2. Model-free evidence

We begin by presenting model-free evidence linking the enhanced transparency introduced by the ICMJE policy to technological novelty. We categorize firms using median policy exposure value into two groups: below- and above-median policy exposure. Next, we split each group temporally into the "pre-policy period" (January 1st, 2000 to September 13th, 2005) and the "post-policy period" (September 14th, 2005 to December 31st, 2014), respectively. We compare the average prior deployments for each sub-group using all trials, as well as early and late-stage trials, where above (below) average prior deployments are associated with less (greater) novelty and lower (higher) risk. We present the results of our analyses in Table 3.

In general, we find that average prior deployments in the post-policy period are higher than the pre-policy period for firms with below- and above-median exposure for all trials (row A), early-stage (row B), and late-stage (row C) trials. This results in a positive difference between average prior deployments in the post- and pre-policy periods (columns 4 and 7). For instance, the positive difference in row A between the post- and pre-period averages of 4.51 for firms with below-median policy exposure (column 4) and 6.23 for firms with above-median policy exposure (column 7) shows that both groups of firms used less novel technologies after the transparency shock. Moreover, the average prior deployments for firms with above-median policy exposure (6.23) is higher compared to firms with below-median policy exposure (4.51), resulting in a positive difference of 1.72 (column 8). This bolsters our contention that firms with greater exposure to the policy were more likely to pursue less novel drug development activities compared to their counterparts.

We replicate the above analysis for sub-samples of trials belonging to early-stage (Phase 1 and 2) and late-stage (Phase 3) trials. The results in rows B and C also yield positive difference-in-differences estimates (column 8), suggesting a disproportionate increase in prior deployments among more exposed firms. In addition, we note that the difference in average prior deployments steadily increases from early-stage to late-stage trials, a pattern consistent with our assertion that more novel technologies are riskier, and therefore less likely to advance through to the late-stage trials. In summary, the results in Table 3 provide model-free evidence of the effect of transparency on technological novelty. Next, we turn to formal econometric analysis to examine our research question.

6.3. Main results

We present the main estimation results from Equation 2 in Table 4. Column 1 presents the estimates of the main effects of Exposure, post-policy indicator, and the interaction term without any control variables. Consistent with the model-free results of Table 3, the difference-in-differences effect estimate is positive and statistically significant. The 0.18 estimate associates one standard deviation higher policy exposure with an average of 19.7% ($e^{0.18} - 1$) increase of prior deployments following the policy's enactment. Given that a percentage increase in prior deployments corresponds to 0.16 years lag (see Section 6.1), a 19.7% increase in the number of prior deployments can also be expressed as firms falling behind the technological frontier by approximately 3.2 years (0.16 years × 19.7%).

Column 2 of Table 4 presents results for the full specification (including all control variables). Again, in this case, the difference-in-differences estimate is positive and statistically significant ($\hat{\beta}$ = 0.13, SE = 0.04). According to this estimate, one standard deviation increase in policy exposure is associated with a 13.9% ($e^{0.13} - 1$) increase in the average prior deployments after the policy's enactment. The effect can also be interpreted as firms falling behind the technological frontier by 2.2 years (0.16 years × 13.9%).

6.4. Robustness checks

6.4.1. Event study and pre-trends

To investigate the effect's temporal unfolding and the potential presence of confounding pre-trends, we estimate an event study version of Equation 2, where we interact Exposure with indicators for successive two-year periods. We present the results of the event study in Figure 4. Each marker in this figure represents the difference-in-differences point estimate for the two-year period ending in the year shown by the horizontal axis. For example, the marker shown for 2001 represents the effect estimated for 2000 and 2001. Similarly, for 2003, the effect is estimated for 2002 and 2003, and so on. The effect for 2004-2005 is normalized to zero.

We see that the estimates for the years 2000-2001 and 2002-2003 are statistically insignificant. As such, these estimates preclude the presence of a confounding pre-trend. The difference-in-differences parameter estimates for the "pre-period" are also jointly insignificant (χ^2= 0.31, p-value = 0.77). The difference-in-differences parameter estimates for the years that follow illustrate that the effect is first perceptible soon after the policy's creation (with 95% confidence), in the 2006-2007 period, and adopts an increasing trend in subsequent years. By the end of the sample (2013-2014), a one standard deviation increase in policy exposure is associated with approximately 21.6% increase in prior deployments which translates to a 3.5-year lag relative to the average technological frontier (0.16 years × 21.6%). These results suggest that the policy's effects on technological novelty may persist and continue to intensify over time.

6.4.2. Falsification

To further probe the causal interpretation of our main estimate of column 2, Table 4, we conduct a falsification test. The test consists of replacing observed firm-level Exposure scores with random values drawn from a standard-normal distribution. Figure 5 shows the distribution of difference-in-differences estimates ($\hat{\beta}$) that we obtain from estimating Equation 2 in 1,000 pseudo-samples created following this procedure. The distribution of estimates is centered at zero and symmetric. As shown by the vertical dashed line, our main estimate from column 2 in Table 3 is located in the right tail of the distribution, yielding an implied p-value of 0.011. That is, the estimate is highly statistically improbable relative to the distribution of estimates obtained via randomization. The falsification results lend confidence to the causal interpretation of our main results.

6.4.3. Alternate measures of technological novelty

Our measure for technological novelty is the number of prior deployments of a technology in clinical trials for drugs other than the focal drug. An implicit assumption in this measure is that each deployment reduces the novelty or risk by the same amount. However, technologies vary greatly and thus the impact of transparency may differ depending on the characteristics of a technology. Therefore, we use three alternate measures of technological novelty and present the robustness of our results in Table 5.

First, we analyze the heterogeneity in the impact of transparency across trials in different phases because they differ in evidentiary requirements and encompass different levels of information. For instance, late-stage Phase 3 trials are distinctly more informative about the prospects of underlying technologies because they rely on randomized-controlled designs and enroll much larger populations when compared to Phase 1 or 2 trials (Van Norman, 2016; Lipsky and Sharp, 2001). Results from late-stage Phase 3 trials receive more attention from investors and the media compared with early-stage Phase 1 or 2 trials. Hence, lackluster late-stage results that are more conspicuous as a result of transparency would cause firms greater distress. Thus, we expect the transparency shock would have hindered novelty more in late-stage than in early-stage trials [Footnote: A series of articles has documented the impacts of negative clinical trial results on firm valuation and performance (Hermosilla, 2021; Hwang, 2013; Girotra et al., 2007; Sharma and Lacey, 2004; Urbig et al., 2013). In line with our argument, this literature finds that: (i) negative trial results are more impactful than the directionally-opposite effects of positive trial results (Hermosilla, 2021; Hwang, 2013; Sharma and Lacey, 2004), and (ii) late-stage negative clinical trial results are more impactful than early-stage ones.]. To examine the effect of transparency by trial phases, we split trials by stages into late-stage (Phase 3) and early-stage (Phase 1 and 2) trials and estimate Equation 2 on the two sub-samples. Consistent with our prediction, column 2 of Table 5 shows that transparency did not have a significant effect on early-stage trials ($\hat{\beta}$ = 0.12, SE = 0.08). In contrast, column 3 shows that the impact of transparency is primarily driven by late-stage trials, as reflected by the positive and statistically significant estimate ($\hat{\beta}$ = 0.16, SE = 0.04).

Second, we test whether firms increase their use of technologies that have resulted in a commercial drug after passing the stage-gate tests. Commercializing a drug generates a significant amount of information, and consequently, the underlying technology could be associated with a greater reduction in risk compared to a drug that is still under development. To incorporate this risk reduction, we construct a version of prior deployments that only records "validated" deployments, i.e., prior deployments of a technology that have led to any other drug reaching the market prior to the initiation of the focal trial. We find that in this sub-sample (Table 5, column 4), the coefficient is positive and statistically significant ($\hat{\beta}$ = 0.17, SE = 0.08). Moreover, it is almost one-third greater in magnitude than the baseline estimate (column 1 of Table 5), implying that firms increase deployments of projects that offer higher risk reduction opportunities.

Third, we explore the "similarity" of a technology as a factor that may influence firms' technological choices. We capture similarity in two ways: prior deployments of the focal technology that targets the same disease condition vs. other conditions and prior deployments of the focal technology that occur within the same firm vs. other firms. We posit that deployments that target the same condition as the focal trial should offer a larger risk reduction when compared to deployments that target different conditions because a firm could leverage its knowledge of condition-specific factors to reduce risk. For instance, a firm could take into account certain types of co-morbidities in the trial and ameliorate potential side effects. Accordingly, we measure prior deployments by counting deployments that target the same condition as the focal trial. The resulting difference-in-differences estimate ($\hat{\beta}$ = 0.14, SE = 0.09) shown in column 5 of Table 5 is positive, significant, and marginally higher than the baseline estimate.

Next, we replicate the above analysis for deployments that were carried out by the same firm as the one sponsoring the focal trial but for any condition (column 6). We assert that the knowledge generated by prior deployments in the same firm is more easily accessible and transferable and entails lower risk when compared to deployments carried out by other firms. We find that the estimate ($\hat{\beta}$ = 0.15, SE = 0.10) is positive but not significant suggesting that the knowledge gleaned from one condition to another may be less useful, even within the same firm. Finally, we operationalize prior deployments of the focal technology at the same firm and in the same condition as of focal trial under consideration (Table 5, column 7). We expect these same-firm, same-targeted condition trials to be informative and risk-reducing to the firm under increased transparency. Our results show that the estimate ($\hat{\beta}$ = 0.24, SE = 0.08) is positive, significant, and considerably larger than all other estimates bolstering our assertion that firms with higher policy exposure increase their prior deployments of "similar" technologies after the policy because they offer better opportunities to reduce risk.

6.5. Mechanism test and other analyses

So far, we have provided causal evidence that firms react to transparency by increasing prior deployments, i.e., shifting investments to relatively less novel technologies. We posit that this may occur to avoid negative information spillovers arising from potential NPD failures (which are harder to hide after transparency). In this section, we investigate this potential mechanism by studying whether the effect varies by a firm's technological diversification. Subsequently, we assess the heterogeneous impact of transparency on exploratory and exploitative R&D.

6.5.1. Technological diversification

In the pharmaceutical industry, negative informational spillovers manifest when the failure of one technology brings into question the viability of other drugs using the same technology or other related technologies (Krieger, 2021; Magazzini et al., 2012). We assert that the relationship between transparency and technological novelty is moderated by a firm's technological diversification due to such negative informational spillover effects. To illustrate our rationale, consider two firms with varying levels of technologically diversified portfolios. Firm A has a portfolio where all the drugs in development rely on the same technology resulting in a technologically concentrated portfolio. A developmental failure in firm A's portfolio is likely to cast a shadow on all the drugs in development because the underlying technology is common to all the drugs. In contrast, firm B has a portfolio in which each of the drugs in development relies on different technologies suggesting a highly diversified portfolio. A developmental failure in firm B's portfolio is not likely to extend to other drugs in development because they use different technologies.

We can generalize this to state that firms with less technologically diversified portfolios are more exposed to risk from trial failure because the failure is likely to affect other trials that rely on the failed technology due to a potential contagion effect. Such firms will be disproportionally affected by the enhanced transparency requirements because they are more vulnerable to negative information spillover. In contrast, firms with more technologically diversified portfolios are able to tolerate trial failure and overcome potential operational and financial setbacks by focusing on trials with other technologies. Therefore, firms with concentrated portfolios will respond to the transparency shock by choosing less novel technologies compared to firms with diversified portfolios. In other words, we expect technological diversification to negatively moderate the impact of transparency on technological novelty. We measure technological diversification with HHI or Herfindahl-Hirschman Index (Garcia-Vega, 2006; Quintana-Garc´ıa and Benavides-Velasco, 2008). For each trial j, we compute the technological diversification metric as:

$$Technological\ Diversification_j = \sum_{t \in \mathcal{T}_j} \left(\frac{|\mathcal{C}_t|}{|\mathcal{P}_j|} \right)^2$$

where, for a trial j sponsored by firm i, we consider all other drugs being tested in clinical trials by the same firm at the time the trial begins. We call this set $\mathcal{P}_j$. Next, we identify the set of all technologies associated with these candidates, $\mathcal{T}_j$, and the set of drugs associated with each technology t, $\{\mathcal{C}_t\}_t \in \mathcal{T}_j$.

Our measure of technological diversification ranges from zero which indicates full diversification (i.e., a different technology for each trial) to one hundred which indicates full concentration (i.e. a single technology underlying every trial). The average, median, and standard deviation of technological diversification are 2.26, 1.35, and 4.07. The low average and median values highlight that the firms in our sample are highly diversified. We also find a positive but statistically insignificant correlation between technological diversification and firm-level exposure values [Footnote: Regression results associate one larger standard deviation of Exposure to 15% larger technological diversification values (p = 0.501). This result does not materially change depending on whether we control or not for temporal trends.]

Next, we categorize our data using the median value of HHI into two sub-samples: below- and above-median HHI which corresponds to high and low levels of technological diversification, respectively. We estimate Equation 2 on the two sub-samples separately and report the difference-indifferences estimates in Table 6. We find that firms with relatively high technological diversification (column 1) show a statistically insignificant effect ($\hat{\beta}$ = −0.10, SE = 0.08), whereas firms with technologically concentrated portfolios (column 2) show a positive and statistically significant effect ($\hat{\beta}$ = 0.18, SE = 0.04). The effect size for firms with technologically concentrated portfolios is about one-third larger than the baseline estimate from the full sample (Table 4, column 2) and indicates that the impact of transparency manifests through firms with low technological diversification.

6.5.2. Additional Analysis: Exploration vs. Exploitation

Next, we demonstrate the implications of the proposed mechanism at the technology level with exploitation-exploration (March, 1991) and ambidexterity lenses (He and Wong, 2004). The exploitation-exploration lens is based on the notion of trading off underlying risks, where exploitation is associated with higher risk compared to exploration, and more exploitation implies less exploration, and vice-versa. Ambidexterity, the alternate paradigm suggests that firms conduct both exploitative and exploratory activities simultaneously, in an effort to balance the underlying risks. It reflects a firm's need for exploitation-oriented R&D which leverages existing knowledge, involves less risk, and has the potential to deliver immediate results, and exploration-oriented R&D which serves as a means to develop future knowledge for sustained competitive advantage, and long-term firm growth and survival. These two perspectives highlight an interesting tension, especially critical in the pharmaceutical industry: does higher exploitation mean reduced exploration?

To examine how transparency affects the exploration of novel technologies and the exploitation of existing technologies, we run three different analyses using different dependent variables and estimation models. First, to assess whether the pro-transparency shock increased the use of exploitative R&D, we compile a sample where prior deployments are greater than 0. Second, to assess whether the pro-transparency shock increased the use of exploratory R&D and novel technologies, we use an indicator variable as a dependent variable which takes the value of 1 if the number of prior deployments is equal to 0, and 0 otherwise. Finally, we assess the impact of the pro-transparency shock on the use of "near-novel" technologies through an indicator variable that equals 1 for trials where the number of prior deployments is greater than zero but less than the average number of prior deployments, and 0 otherwise. We describe the empirical approach used to fit Equation 2 for each of the sub-samples below and present the results in Table 7.

To assess whether the pro-transparency shock increased exploitation, we use a Poisson quasi- maximum likelihood model to fit Equation 2 on the sample of trials associated with strictly positive values of prior deployments. The estimate is positive and significant (column 1) and indicates that the effect is driven by an emphasis on previously-deployed technologies, i.e., by an increase in the use of existing technologies. Next, we estimate the impact of transparency on the likelihood of exploring purely novel technologies. Since our dependent variable is a binary variable, we use a logistics regression model to fit Equation 2. The results (Table 7, column 2) show that the difference-in-differences coefficient is statistically insignificant. This suggests that transparency did not impact a firm's exploratory activities. Specifically, transparency neither inhibited nor increased the exploration of purely novel technologies.

Finally, to assess the impact of transparency on the likelihood of firms using near-novel technologies, we use the logistic regression to fit Equation 2 on the sub-sample of trials without pure exploratory trials, i.e., we include all trials with strictly positive prior deployments in the subsample. Interestingly, the difference-in-differences coefficient is negative and statistically significant (column 3), indicating that firms with higher exposure to the policy are less likely to use a technology that has not been used enough. Specifically, the effect size of −0.19 amounts to a 17.3% reduction in near-novel exploratory efforts due to the policy. We also test the moderation effect of diversification in this sub-sample by replicating the analysis in Section 6.5.1. The results reiterate that the reduction in exploratory activities is driven by firms with highly diversified technology portfolios. As a set, these results provide strong evidence that while the increase in transparency leads firms to exploit previously used technologies, it may also have affected their exploration of near-novel technologies.

7. DISCUSSION AND CONCLUSION

Our study investigates the impact of transparency in NPD efforts on technological novelty. Using a shock that increased transparency in the pharmaceutical sector, we show that transparency negatively impacts technological novelty. We reason that the increased transparency makes it harder for firms to hide a potential NPD failure and therefore, firms shy away from pursuing the search for novel technologies as a means to reduce their vulnerability to negative informational spillovers. As a response, firms increase the use of technologies that they have used before, leading to a 13.9% reduction in technological novelty due to transparency, which corresponds to a 2.2-year lag from the average technological frontier. The results are robust to various measures and falsifications. Further, we show that the effect varies by firms' technological diversification - while firms with more diversified portfolios are unaffected by the transparency shock, those with less diversified portfolios increase their reliance on previously used technologies to avert negative information spillovers of the failed technology.

One of our main findings highlights the nuanced role of ambidexterity in R&D. While the exploitation-exploration lens would have suggested that an increase in exploitation would result in lower exploration, we find that although the transparency shock leads to higher exploitation, firms do not decrease “first-time” use of novel technologies, instead, they reduce the use of technologies that have been used less often. Firms may continue to invest in purely novel technologies because the independent and simultaneous pursuit of exploratory and exploitative activities is important for firm survival and growth and require orthogonal skills and organizational resources (March, 1991; Chandrasekaran et al., 2015). However, a concurrent reduction in near-novel exploratory efforts highlights a need for a more nuanced understanding of ambidexterity in the NPD context, by going from exploitation-exploration classification to a more granular view that incorporates “near-novel” exploration. These results complement findings in the innovation diffusion literature. It is likely that transparency enables a faster diffusion of publicly available trial information. This could lead to an increase in environmental instability driven by an increase in the rate of change of technology (Henderson and Clark, 1990; Utterback, 1994). Chao and Kavadias (2008) show that under high environmental instability, firms reduce risk by increasing the number of incremental R&D projects and not by reducing the number of radical R&D projects.

Our results are pertinent beyond the pharmaceutical industry, especially in light of recent scandals that have intensified calls for business transparency in other industries. Consider an example from the commercial aviation industry. Boeing, one of the largest airplane manufacturers, concealed crucial information about the known risks associated with their new aircraft 737 Max. After two plane crashes and 346 lost lives, the Department of Justice charged Boeing with criminal charges related to defrauding the Federal Aviation Administration’s Aircraft Evaluation Group (US Department of Justice, 2021). Theranos - the now-defunct silicon valley start-up is another example. It claimed to have developed a revolutionary diagnostic blood test. A Wall Street Journal investigation later revealed that the company knowingly pedaled fraudulent claims using inaccurate test results (Carreyrou, 2015). These firms did not disclose negative NPD outcomes (higher failure risk in the case of Boeing or lack of benefits in the case of Theranos) because such disclosures were not aligned with the firms’ business interests. Such proclivity of firms towards NPD secrecy compels regulatory bodies and industry watchdogs to enforce transparency measures. By accentuating the unintended outcomes of transparency, our results underscore the importance of appropriately managing transparency initiatives for more novel innovation.

7.1. Implications for managers

Practitioners and scholars note that innovation efforts within the drug industry have plateaued resulting in the lack of novel drugs in the market. The past editor-in-chief of the New England Journal of Medicine asserted that pharmaceutical firms are more inclined to produce “me-too” versions of their successful drugs (Angell, 2021). The declining number of patents involving new chemical entities (Ivanenkov et al., 2019) coupled with the increasing number of commercial drugs based on a limited set of tried and tested therapeutic mechanisms (Shih et al., 2018) have resulted in a novelty deficit in the drug industry (Gibney, 2022). We provide evidence that increasing transparency measures might have contributed to the novelty crisis in the pharmaceutical industry by amplifying the use of existing technologies. Risk aversion has been shown to keep large pharmaceutical firms from exploring novel R&D search (Krieger et al., 2022). We show that transparency, too, could exacerbate risk aversion among large pharmaceutical firms leading to a cutback in novel pursuits.

Risk aversion has implications for collaborations within the pharmaceutical industry. Specifically, increased transparency and the associated risk aversion could explain the growing trend of outsourcing pharmaceutical R&D where more than 50% of the new drugs are brought to market by specialized startup firms or academic research centers (Kneller, 2010; Rafols et al., 2014). Collaborating with academic research centers and specialized start-up firms to perform their early-stage research allows large pharmaceutical firms to transfer early risks in the drug discovery process to the partners. It also provides them the opportunity to acquire/in-license the less risky, late-stage drug development from those partners (Angell, 2021). Anecdotal evidence suggests that large pharmaceutical firms leverage such partnership models to avoid patent cliffs in the past (Gibney, 2022). Transparency measures, too, could have contributed to the growth of outsourcing high-risk, early-stage drug discovery because it allows firms to avoid the potential negative information spillovers arising from increased transparency.

Lastly, the impact of transparency on the exploration-exploitation trade-off is of particular managerial concern because it is consequential for long-term performance. Prior research suggests that a firm needs to strike a balance between incremental and radical projects (Cooper et al., 1997; Mudambi and Swift, 2014). We show this balance could be disrupted as firms react to increased transparency by changing their NPD approach. A shift toward more exploitation and less exploration could result in underperformance (Chao et al., 2009; He and Wong, 2004) in the best case. It could be self-destructive (March, 1991) and detrimental to the firm's survival (Swift, 2016) in the worst case.

One way that managers could mitigate the impact of transparency is by maintaining a diversified portfolio. We show that the unintended consequences of transparency are most prominent in firms with concentrated portfolios, and transparency does not impact firms with diversified portfolios. Diversification could curb negative information spillovers and reduce the higher failure costs associated with increased transparency. While specialization has been highlighted as a key strength that leads to better learning and productivity (Narayanan et al., 2009), firms with high dependence on a small set of technologies could be more vulnerable to environmental shocks. Our study adds to the literature that highlights the benefit of diversification in innovation management (Klingebiel and Rammer, 2014).

The moderating effect of diversification on the impact of transparency is particularly important in our context because pharmaceutical firms are increasingly facing higher pressure to be transparent. Moreover, many firms have publicly committed to "open innovation" and promised to be transparent about clinical trials and their results (Hunter and Stephens, 2010). Furthermore, as discussed earlier, there is a growing trend of novel drugs coming from firms with more specialized portfolios. The increasing need for transparency coupled with the high reliance on R&D originating in firms with specialized technological portfolios bolsters the importance of diversification in limiting the negative impact of transparency on technological novelty.

7.2. Implications for policymakers

Our results highlight an important trade-off between transparency and technological novelty. This trade-off is more critical in the context of pharmaceutical NPD. While clinical trial transparency helps solve the issue of selective reporting of trial results thereby safeguarding patients' interests, it negatively impacts technological novelty. At an industry level, the negative effect of transparency on innovation may impact for patient populace awaiting breakthrough drugs. Prior literature documents evidence of pharmaceutical firms pursuing "incremental innovation" – prioritizing small improvements in their existing R&D portfolio over pursuing novel breakthrough innovation (Munos and Chin, 2011). We show that transparency could exacerbate incremental innovation in that it leads to higher exploitation of previously used technologies and could lead to lower exploration of novel technologies.

Our results indicate that to offset the unintended consequences of transparency, policymakers and regulators such as the FDA may need to consider financial incentives and policy solutions to spur novel innovation. The FDA has made significant strides in this direction by providing faster review times under its special drug approval programs such as breakthrough therapy designation, accelerated approval, fast-track programs, etc. but more could be done. Regulatory authorities could implement programs that reward novel innovations differentially. For instance, regulatory authorities can institute policies that extend the patent duration or grant longer market exclusivity for more novel drugs to incentivize firms to take greater R&D risks. They can also provide financial support to firms that examine certain types of drugs such as those with orphan drug designation. Additionally, to reduce long development times and uncertain outcomes, FDA can provide faster turnaround time, especially for drugs that show early therapeutic promise.

7.3. Limitations and future research

An important caveat about our analysis is that it is circumscribed to a segment of the drug development sector: large pharmaceutical firms. While we expect that transparency may also affect the innovation decisions of smaller firms (e.g., biotech startups), we suspect that the mechanisms at play may be different. Perhaps most prominently, these differences may stem from the size of technological choice sets. Because large firms have broad portfolios and the ability to in-license, they have ample scope to choose "the right technology" to develop, In contrast, a typical biotech startup firm is centered around developing a specific technology, and may lack the technological choices that large firms have. Accordingly, if transparency affects innovation by small firms, it may do so via entry/exit rather than through portfolio composition. Financing structures (public investment markets for large firms, Venture Capital for small ones) and the role of academic publishing introduce further differences that we think may be important. From a practical standpoint, the ICMJE policy may be inadequate to study how smaller biotech firms react to increased transparency initiatives, especially given their high turnover rates, few firms may have survived the entire study duration. Perhaps, future researchers could examine the impact of transparency initiatives more appropriate and relevant to small firms.

Second, although our study shows an unintended, negative outcome of increased transparency on technological novelty, we acknowledge the potential benefits of transparency. The ICMJE policy was implemented to address problems associated with selective reporting of the clinical trial results in the pharmaceutical industry and could have led to benefits for key stakeholders including patients, firms, and providers. For instance, the increase in trial registration following the ICMJE policy could have instilled greater confidence among trial subjects resulting in an increase in trial participation. For pharmaceutical firms, increased transparency might have deterred firms from hiding or misreporting trial outcomes (such as Paxil) leading to a reduction in the number of corporate violations and misdemeanors. By highlighting trial results that show poor safety or efficacy of trial drugs, the policy could have helped physicians make a more informed decision in prescribing drugs. Future researchers could quantify the benefits of transparency by examining these relationships. Additionally, researchers can also examine other negative consequences of transparency, and help generalize our findings.

Third, our analysis with exploration as novel and near-novel hints at the possibility that firms consider novelty to be a continuous spectrum instead of a dichotomous construct. We invite future researchers to explicitly test this idea. The questions around how firms determine technological novelty – the very first use, less than average use, or some other measure, and their repercussions on NPD decisions are an interesting line of inquiry for future research.

REFERENCES

Acemoglu, D., D. H. Autor, and D. Lyle (2004). Women, war, and wages: The effect of female labor supply on the wage structure at midcentury. Journal of political Economy 112 (3), 497–551. 12

Angell, M. (2021, Aug). Big pharma, bad medicine. 23, 24 Arora, A., S. Belenzon, and A. Patacconi (2018). The decline of science in corporate r&d. Strategic Management Journal 39 (1), 3–32. 14

Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? Journal of Economics & Management Strategy 11 (4), 551–594. 8, 15

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. Journal of applied econometrics 18 (1), 1–22. 10
Beer, R., I. Rios, and D. Saban (2021). Increased transparency in procurement: The role of peer effects. Management Science 67 (12), 7511–7534. 3, 5, 33
Benner, M. J. and M. L. Tushman (2003). Exploitation, exploration, and process management: The productivity dilemma revisited. Academy of management review 28 (2), 238–256. 6
Bernstein, E. S. (2017). Making transparency transparent: The evolution of observation in management theory. Academy of Management Annals 11 (1), 217–266. 5
Bhaskarabhatla, A. and D. Hegde (2014). An organizational perspective on patenting and open innovation. Organization Science 25 (6), 1744–1763. 15
Brown, J. R. and G. Martinsson (2019). Does transparency stifle or facilitate innovation? Management Science 65 (4), 1600–1623. 5
Brownlee, S. (2004). Drug-trial registry, the. The New York Times. Available at https://www.nytimes.com/2004/12/12/magazine/drugtrial-registry-the.html 9
Buell, R. W. and B. Kalkanci (2021). How transparency into internal and external responsibility initiatives influences consumer choice. Management Science 67 (2), 932–950. 5
Buell, R. W., T. Kim, and C.-J. Tsay (2017). Creating reciprocal value through operational transparency. Management Science 63 (6), 1673–1695. 5, 33
Buell, R. W. and M. I. Norton (2011). The labor illusion: How operational transparency increases perceived value. Management Science 57 (9), 1564–1579. 5, 33
Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. American Economic Review 102 (6), 2981–3003. 3, 5, 33
Carreyrou, J. (2015, October). Hot startup theranos has struggled with its blood-test technology. 23
Chan, T., J. A. Nickerson, and H. Owan (2007). Strategic management of r&d pipelines with cospecialized investments and technology markets. Management Science 53 (4), 667–682. 11
Chandrasekaran, A., K. Linderman, and R. Schroeder (2015). The role of project and organizational context in managing high-tech r&d projects. Production and Operations Management 24 (4), 560–586. 6, 23
Chandy, R. K. and G. J. Tellis (2000). The incumbent's curse? incumbency, size, and radical product innovation. Journal of marketing 64 (3), 1–17. 6
Chao, M. and I. Larkin (2022). Regulating conflicts of interest in medicine through public disclosure: Evidence from a physician payments sunshine law. Management Science 68 (2), 1078–1094. 5
Chao, R. O. and S. Kavadias (2008). A theoretical framework for managing the new product development portfolio: When and how to use strategic buckets. Management science 54 (5), 907–921. 6, 23
Chao, R. O., S. Kavadias, and C. Gaimon (2009). Revenue driven resource allocation: Funding authority, incentives, and new product development portfolio management. Management Science 55 (9), 1556–1569.24
Cheng, Y.-T. and A. H. Van de Ven (1996). Learning the innovation journey: order out of chaos? Organization science 7 (6), 593–614. 6
Cooper, R. G. (1990). Stage-gate systems: a new tool for managing new products. Business horizons 33 (3), 44–54. 7
Cooper, R. G., S. J. Edgett, and E. J. Kleinschmidt (1997). Portfolio management in new product development: Lessons from the leaders—ii. Research-Technology Management 40 (6), 43–52. 24
Cullen, Z. and R. Perez-Truglia (2022). How much does your boss make? the effects of salary comparisons. Journal of Political Economy 130 (3), 766–822. 5, 33
De Angelis, C., J. M. Drazen, F. A. Frizelle, C. Haug, J. Hoey, R. Horton, S. Kotzin, C. Laine, A. Marusic, A. J. P. Overbeke, et al. (2004). Clinical trial registration: a statement from the

international committee of medical journal editors. Annals of internal medicine 141 (6), 477–478. 2, 9
Ding, M. and J. Eliashberg (2002). Structuring the new product development pipeline. Management Science 48 (3), 343–363. 11
Dowden, H. and J. Munro (2019). Trends in clinical success rates and therapeutic focus. Nat. Rev. Drug Discov 18 (7), 495–6. 7
Dranove, D., C. Garthwaite, and M. Hermosilla (2022). Does consumer demand pull scientifically novel drug innovation? The RAND Journal of Economics 53 (3), 590–638. 1, 7, 10
Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite (2003). Is more information better? the effects of "report cards" on health care providers. Journal of political Economy 111 (3), 555–588. 5, 33
Duggan, M. and F. S. Morton (2010). The effect of medicare part d on pharmaceutical prices and utilization. American Economic Review 100 (1), 590–607. 12
Eggers, J. and A. Kaul (2018). Motivation and ability? a behavioral perspective on the pursuit of radical invention in multi-technology incumbents. Academy of Management Journal 61 (1), 67–93. 6
Gans, J. S., F. E. Murray, and S. Stern (2017). Contracting over the disclosure of scientific knowledge: Intellectual property and academic publication. Research Policy 46 (4), 820–835. 8
Garcia-Vega, M. (2006). Does technological diversification promote innovation?: An empirical analysis for european firms. Research policy 35 (2), 230–246. 20
Gibney, M. (2022, Apr). The 'innovation gap' haunts big pharma, and the race is on for the next blockbusters. 23, 24
Gibson, C. B. and J. Birkinshaw (2004). The antecedents, consequences, and mediating role of organizational ambidexterity. Academy of management Journal 47 (2), 209–226. 6
Girotra, K., C. Terwiesch, and K. T. Ulrich (2007). Valuing r&d projects in a portfolio: Evidence from the pharmaceutical industry. Management Science 53 (9), 1452–1466. 8, 11, 18
Greenberg, M. D. (1999). Aids, experimental drug approval, and the fda new drug screening process. NYUJ Legis. & Pub. Pol'y 3, 295. 8
Guo, T., S. Sriram, and P. Manchanda (2021). The effect of information disclosure on industry payments to physicians. Journal of Marketing Research 58 (1), 115–140. 33
Hara, T. (2003). Innovation in the pharmaceutical industry: the process of drug discovery and development. Edward Elgar Publishing. 7
He, Z.-L. and P.-K. Wong (2004). Exploration vs. exploitation: An empirical test of the ambidexterity hypothesis. Organization science 15 (4), 481–494. 3, 21, 24
Henderson, R. and I. Cockburn (1994). Measuring competence? exploring firm effects in pharmaceutical research. Strategic management journal 15 (S1), 63–84. 8
Henderson, R. M. and K. B. Clark (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. Administrative science quarterly, 9–30. 23
Hermosilla, M. (2021). Rushed innovation: Evidence from drug licensing. Management Science 67 (1), 257–278. 10, 18
Hermosilla, M. (2022). Ethical experimentation and regulatory flexibility in drug development. Working Paper, Johns Hopkins University. 10
Hermosilla, M., F. Gutierrez-Navratil, and J. Prieto-Rodriguez (2018). Can emerging markets tilt global product design? impacts of chinese colorism on hollywood castings. Marketing Science 37 (3), 356–381. 12
Hunter, J. and S. Stephens (2010). Is open innovation the way forward for big pharma? Nature Reviews Drug Discovery 9 (2), 87–88. 24
Hwang, T. J. (2013). Stock market returns and clinical trial results of investigational compounds: an event study analysis of large biopharmaceutical companies. PloS one 8 (8), e71966. 18

Ivanenkov, Y. A., B. A. Zagribelnyy, and V. A. Aladinskiy (2019). Are we opening the door to a new era of medicinal chemistry or being collapsed to a chemical singularity? perspective. Journal of medicinal chemistry 62 (22), 10026–10043. 23
Jin, G. Z. and P. Leslie (2003). The effect of information on product quality: Evidence from restaurant hygiene grade cards. The Quarterly Journal of Economics 118 (2), 409–451. 33
Juliano, R. (2013). Pharmaceutical innovation and public policy: The case for a new strategy for drug discovery and development. Science and Public Policy 40 (3), 393–405. 1
Kahneman, D. and D. Lovallo (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. Management science 39 (1), 17–31. 1
Kahneman, D. and A. Tversky (1973). On the psychology of prediction. Psychological review 80 (4), 237. 1
Kesselheim, A. S. and M. M. Mello (2007). Confidentiality laws and secrecy in medical research: improving public access to data on drug safety. Health Affairs 26 (2), 483–491. 8
Khmelnitskaya, E. (2021). Competition and Attrition in Drug Development. Ph. D. thesis, University of Virginia. 10
Kirchhoff, B. A. (1992). Entrepreneurship's contribution to economics. Entrepreneurship theory and practice 16 (2), 93–112. 6
Klingebiel, R. and C. Rammer (2014). Resource allocation strategy for innovation portfolio management. Strategic Management Journal 35 (2), 246–268. 24
Kneller, R. (2010). The importance of new companies for drug discovery: origins of a decade of new drugs. Nature reviews Drug discovery 9 (11), 867–882. 24
Kondro, W. (2004). Drug company experts advised staff to withhold data about ssri use in children. CMAJ 170 (5), 783–783. 9
Kraft, T., L. Vald´es, and Y. Zheng (2018). Supply chain visibility and social responsibility: Investigating consumers' behaviors and motives. Manufacturing & Service Operations Management 20 (4), 617–636. 5, 33
Krieger, J., D. Li, and D. Papanikolaou (2022). Missing novelty in drug development. The Review of Financial Studies 35 (2), 636–679. 4, 7, 10, 24
Krieger, J. L. (2021). Trials and terminations: Learning from competitors' r&d failures. Management Science 67 (9), 5525–5548. 2, 8, 10, 20
Lee, J., H. S. Lee, H. Shin, and V. Krishnan (2021). Alleviating drug shortages: The role of mandated reporting induced operational transparency. Management Science 67 (4), 2326–2339. 5, 33
Lerner, J. and U. Malmendier (2010). Contractibility and the design of research agreements. American Economic Review 100 (1), 214–46. 15
Lipsky, M. S. and L. K. Sharp (2001). From idea to market: the drug approval process. The Journal of the American Board of Family Practice 14 (5), 362–367. 18
Lo, A. W. and R. T. Thakor (2022). Financing biomedical innovation. Annual Review of Financial Economics 14, 231–270. 4
Lovallo, D., T. Koller, R. Uhlaner, and D. Kahneman (2020). Your company is too risk-averse. Harvard Business Review 98 (2), 104–111. 1
Lowe, D. (2015). Target invalidation. In the Pipeline. Available at https://www.science.org/content/blog-post/target-invalidation. 8
Magazzini, L., F. Pammolli, and M. Riccaboni (2012). Learning from failures or failing to learn? Lessons from pharmaceutical r&d. European Management Review 9 (1), 45–58. 2, 8, 20
March, J. G. (1991). Exploration and exploitation in organizational learning. Organization science 2 (1), 71–87. 6, 21, 23, 24
Mas, A. and E. Moretti (2009). Peers at work. American Economic Review 99 (1), 112–45. 5, 33
Meier, B. (2004a). Group weighs plan for full drug-trial disclosure. The New York Times. Available at https://www.nytimes.com/2004/06/15/business/group-weighs-plan-for-full-drug-trial-disclosure.html . 9

Meier, B. (2004b). Major medical journals will require registration of trials. The New York Times. Available at https://www.nytimes.com/2004/09/09/business/major-medical-journals-will-require-registration-of-trials.html. 9

Mejia, J., G. Urrea, and A. J. Pedraza-Martinez (2019). Operational transparency on crowdfunding platforms: effect on donations for emergency response. Production and Operations Management 28 (7), 1773–1791. 5, 33

Miller, J. E., M. Wilenzick, N. Ritcey, J. S. Ross, and M. M. Mello (2017). Measuring clinical trial transparency: an empirical analysis of newly approved drugs and large pharmaceutical companies. BMJ open 7 (12), e017917. 1

Mohan, B., R. W. Buell, and L. K. John (2019). Lifting the veil: The benefits of cost transparency. Harvard Business School NOM Unit Working Paper (15-017), 15–017. 5

Mudambi, R. and T. Swift (2014). Knowing when to leap: Transitioning between exploitative and explorative r&d. Strategic Management Journal 35 (1), 126–145. 6, 24

Munos, B. H. and W. W. Chin (2011). How to revive breakthrough innovation in the pharmaceutical industry. Science translational medicine 3 (89), 89cm16–89cm16. 1, 7, 25

Narayanan, S., S. Balasubramanian, and J. M. Swaminathan (2009). A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. Management science 55 (11), 1861–1876. 24

Parker, C., K. Ramdas, and N. Savva (2016). Is it enough? evidence from a natural experiment in india's agriculture markets. Management Science 62 (9), 2481–2503. 12

Peng, D. X., R. G. Schroeder, and R. Shah (2008). Linking routines to operations capabilities: A new perspective. Journal of operations management 26 (6), 730–748. 5

Polidoro, F. and M. Theeke (2012). Getting competition down to a science: The effects of technological competition on firms' scientific publications. Organization Science 23 (4), 1135–1153. 8, 15

Quintana-Garc´ıa, C. and C. A. Benavides-Velasco (2008). Innovative competence, exploration and exploitation: The influence of technological diversification. Research policy 37 (3), 492–507. 20

Rafols, I., M. M. Hopkins, J. Hoekman, J. Siepel, A. O'Hare, A. Perianes-Rodr´ıguez, and P. Nightingale (2014). Big pharma, little science?: A bibliometric perspective on big pharma's r&d decline. Technological forecasting and social change 81, 22–38. 24

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological bulletin 86 (3), 638. 8

Rothaermel, F. T. and D. L. Deeds (2004). Exploration and exploitation alliances in biotechnology: A system of new product development. Strategic management journal 25 (3), 201–221. 6

Sataline, S. (2005). Medical journal: Drug studies hide key data. The Wall Street Journal. Available at https://www.wsj.com/articles/SB113582179697133550. 9

Schiermeier, Q. (2010). The way back. Nature 466 (7304), 402–403. 8

Schumpeter, J. A. (2013). Capitalism, socialism and democracy. routledge. 6

Sharma, A. and N. Lacey (2004). Linking product development outcomes to market valuation of the firm:
The case of the us pharmaceutical industry. Journal of Product Innovation Management 21 (5), 297–308. 18

Shih, H.-P., X. Zhang, and A. M. Aronov (2018). Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications. Nature Reviews Drug Discovery 17 (1), 19–33. 23

Stern, S. (2004). Do scientists pay to be scientists? Management science 50 (6), 835–853. 8

Subramanian, A. M., M. L´evesque, and V. Van De Vrande (2020). "pulling the plug:" time allocation between drug discovery and development projects. Production and Operations Management 29 (12), 2851–2876. 10

Swift, T. (2016). The perilous leap between exploration and exploitation. Strategic Management Journal 37 (8), 1688–1698. 6, 24

Tversky, A. and D. Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty 5 (4), 297–323. 1

Uotila, J., M. Maula, T. Keil, and S. A. Zahra (2009). Exploration, exploitation, and financial performance: Analysis of s&p 500 corporations. Strategic management journal 30 (2), 221–231. 6

Urbig, D., R. B¨urger, H. Patzelt, and L. Schweizer (2013). Investor reactions to new product development failures: The moderating role of product development stage. Journal of Management 39 (4), 985–1015. 18

US Department of Justice, D. (2021, January). Boeing charged with 737 max fraud conspiracy and agrees to pay over usd 2.5 billion. 23

Utterback, J. M. (1994). Radical innovation and corporate regeneration. Research Technology Management 37 (4), 10. 23

Van Norman, G. A. (2016). Drugs, devices, and the fda: part 1: an overview of approval processes for drugs. JACC: Basic to Translational Science 1 (3), 170–179. 18

Vangamudi, B., T. A. Paul, P. K. Shah, M. Kost-Alimova, L. Nottebaum, X. Shi, Y. Zhan, E. Leo, H. S. Mahadeshwar, A. Protopopov, et al. (2015). The smarca2/4 atpase domain surpasses the bromodomain as a drug target in swi/snf-mutant cancers: insights from cdna rescue and pfi-3 inhibitor studies. Cancer research 75 (18), 3865–3878. 8

Viscusi, W. K., W. A. Magat, and J. Huber (1987). An investigation of the rationality of consumer valuations of multiple health risks. The RAND journal of economics, 465–479. 1

Zarin, D. A., T. Tse, and N. C. Ide (2005). Trial registration at clinicaltrials. gov between may and October 2005. New England Journal of Medicine 353 (26), 2779–2787. 2, 9

Zarin, D. A., T. Tse, R. J. Williams, and T. Rajakannan (2017). Update on trial registration 11 years after the icmje policy was established. New England Journal of Medicine 376 (4), 383–391. 9

Zucker, L. G. and M. R. Darby (1997). Present at the biotechnological revolution: transformation of technological identity for a large incumbent pharmaceutical firm. Research Policy 26 (4-5), 429–446. 8, 15

Table 1: Review of literature on transparency

Article	Unit of analysis	Assessment of Transparency	Performance outcome	Relevant results
Dranove et al. (2003)	Individual patient & hospitals	Disclosing clinical performance of hospitals & physicians (I)	Healthcare demand; Health outcome; Patient selection	The number of cardiac surgeries increased; the proportion of sicker patients getting the surgery decreased (U; U)
Jin and Leslie (2003)	Restaurant	Public display of hygiene inspection grades (I)	Inspection scores; Demand patterns; Foodborne illness Hospitalizations	Health inspection scores increased (D; I) The sensitivity of demand with hygiene quality increased; foodborne illness hospitalizations decreased (D; U)
Mas & Moretti (2009)	Individual employee	Observability of a worker's effort by peers with high relative productivity (V)	Productivity	Increased productivity for low-productivity employees; no decrease in productivity for high-productivity employees (D; I)
Buell & Norton (2011)	Individual customer	Display of effort in work to the customer (V)	Perceived service value; Customer satisfaction; Repurchase Intentions	Increase in customer's perceived value of the service, satisfaction, & intent to repurchase the service (D; I)
Card et al. (2012)	Individual employee	Ability to see the annual salary of peers (I)	Job satisfaction; Job search intention	Decrease in satisfaction & increase in job search intentions for below-median wage earners; no impact on above-median wage-earners (U; U)
Buell et al. (2017)	Individual customer & employee	Customer's ability to see employees/process; Employee's ability to see customers (V)	Employee satisfaction; Service quality; Service efficiency	Customer seeing process: Increase in perceived quality & decreased throughput time; Employee seeing customer: Increase in job satisfaction of the employee (D; I)
Kraft et al. (2018)	Individual customer	Communicating social responsibility practices to consumers (V)	Increase in customer's product valuation	Customer's valuation for the products (D; I)
Mejia et al. (2019)	Individual campaign	Updates on the progress a crowdfunding campaign makes; Showing the certification of a charity (V)	Donation amount	Increase in donation amount when work updates & certification were provided to donors (D; I)
Guo et al. (2021)	Individual physician	Public disclosure of payments physicians receive from pharmaceutical firms (I)	Magnitude of Payments	The magnitude of payments associated with more (less) expensive drugs & heavier (lighter) prescribers increased (decreased) (U; U)
Beer et al. (2021)	Individual employee	Observability of the previous purchasing decision (I)	Cheap/expensive Supplier	Increase in the likelihood of choosing an expensive supplier when a peer did so; the likelihood did not change if the peer chose a cheaper option (U; U)

Lee et al. (2021)	Firm	Reporting of manufacturing interruptions likely to cause shortages (I)	Time to recovery; Annual days of shortage	Decrease in time to recovery & days of shortage; the effect more prominent in presence of competition (D; I)
Cullen & Perez-Truglia (2022)	Individual employee	Visibility of managers' or peers' salary (I)	Hours spent in office; Number of emails sent; Sales	Decrease (increase) in employee effort & the number of emails with peers' (managers') salary visibility (U; U)

* - V=voluntary initiative; I=involuntary initiative. # - D= desirable; U = undesirable; I = intended; U = unintended.

Table 2: Descriptive statistics

	Time period	Minimum	Maximum	Mean	Median	Std. Dev.
(1) Number of trials	2000-2014	87.0	1,205.0	422.7	300.0	330.5
(2) Number of publications	2000-2005	11.0	1,118.0	309.1	292.0	293.0
(3) Number of drugs	2000-2005	3.0	185.0	42.8	24.0	47.0
(4) Raw exposure	2000-2005	0.8	38.4	8.9	8.7	8.0
(5) Exposure	2000-2005	−1.0	3.7	0.0	0.0	1.0
(6) Technological Novelty	2000-2014	0.0	124.0	16.3	10.0	18.8

Notes. Descriptive statistics for all variables are reported at a firm level except technological novelty which is reported at a trial level. The period 2000-2005 corresponds to the period from January 1st, 2000 to September 13th, 2005 which is the pre-policy period. The period 2000-2014 corresponds to the period from January 1st, 2000 to December 31st, 2014 which is the total study duration.

Table 3: Model-free difference-in-difference evidence.

Average prior deployments	Below-median Exposure			Above-median Exposure			Difference-in-differences
	Pre	Post	Post − Pre	Pre	Post	Post − Pre	
(1)	(2)	(3)	(4)=(3)−(2)	(5)	(6)	(7)=(6)−(5)	(8)=(7)−(4)
(A) All trials	13.98	18.49	4.51	11.90	18.13	6.23	1.72
(B) Early-stage trials	12.94	15.94	3.00	11.45	15.90	4.45	1.45
(C) Late-stage trials	18.00	25.03	7.03	13.74	25.33	11.59	4.56

Notes. Higher values of prior deployments reflect less novel new product development. The pre-period corresponds from January 1st, 2000 to September 13th, 2005 while the post-period corresponds from September 14th, 2005 to December 31st, 2014. Early-stage trials belong to either Phase 1 or Phase 2 trials. Late-stage trials are Phase 3 trials. Positive difference-in-differences in the last column indicate that firms that are more exposed to the policy disproportionately increase prior deployments more than firms that are less exposed to the policy.

Table 4: Poisson quasi-maximum likelihood difference-in-differences estimation with prior deployments as the dependent variable.

	(1)	(2)
Post	0.27***	
	(0.06)	
Exposure	−0.04	
	(0.05)	
Exposure×Post	0.18**	0.13**
	(0.08)	(0.06)
Sample	All trials	All trials
N	9,723	9,723
Biologic	N	Y
Year fixed effects	N	Y
Therapeutic area fixed effects	N	Y
Firm fixed effects	N	Y
Phase fixed effects	N	Y

Notes. Robust standard errors (clustered at the firm level) are presented in parentheses. Legend: $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$. We use Poisson quasi-maximum likelihood estimation because our dependent variable is an overdispersed count variable.

Table 5: Poisson quasi-maximum likelihood difference-in-differences estimation with alternate measures of prior deployments as the dependent variables.

	Baseline	Trial phase		Validity	Similarity		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Exposure×Post	0.13**	0.12	0.16***	0.17**	0.14**	0.15	0.24***
	(0.06)	(0.08)	(0.04)	(0.08)	(0.07)	(0.10)	(0.08)
	Prior deployment						
Sponsor	Any firm	Any firm	Any firm	Any firm	Any firm	Same firm	Same firm
Target	Any cond.	Any cond.	Any cond.	Any cond.	Same cond.	Any cond.	Same cond.
Reach	Any phase	Any phase	Any phase	Market	Any phase	Any phase	Any phase
Sample	All trials	Early-stage trials	Late-stage trials	All trials	All trials	All trials	All trials
N	9,723	7,357	2,366	9,723	9,723	9,723	9,723
Biologic	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Therapeutic area fixed effects	Y	Y	Y	Y	Y	Y	Y
Firm fixed effects	Y	Y	Y	Y	Y	Y	Y
Phase fixed effects	Y	Y	Y	Y	Y	Y	Y

Notes. Robust standard errors (clustered at the firm level) are presented in parentheses. Legend: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. We use Poisson quasi-maximum likelihood estimation because our dependent variable is an overdispersed count variable.

Table 6: Poisson quasi-maximum likelihood difference-in-differences estimation on sub-samples split by HHI with prior deployments as the dependent variable.

	(1)	(2)
Exposure×Post	−0.10	0.18***
	(0.08)	(0.04)
Sample	Below-median HHI = High technological diversification	Above-median HHI = Low technological diversification
N	4,862	4,861
Biologic	Y	Y
Year fixed effects	Y	Y
Therapeutic area fixed effects	Y	Y
Firm fixed effects	Y	Y
Phase fixed effects	Y	Y

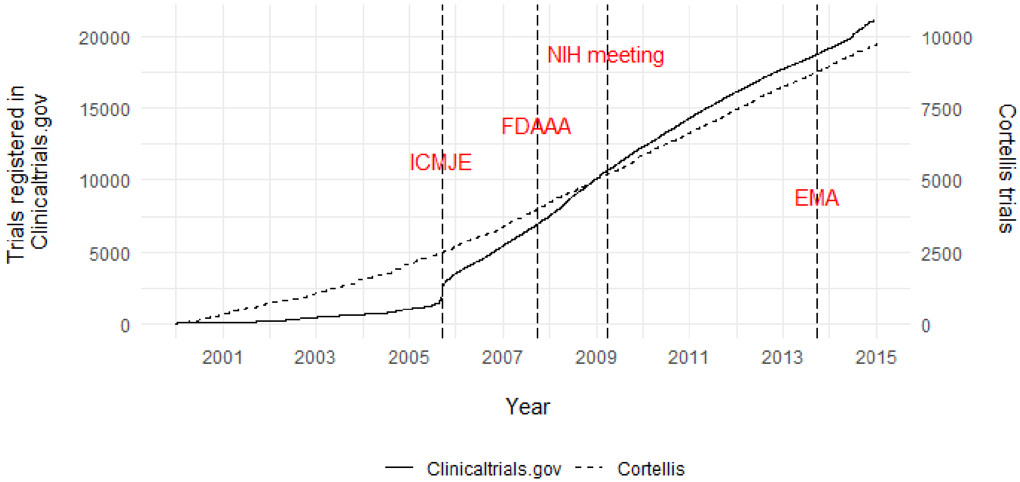
Notes. Robust standard errors (clustered at the firm level) are presented in parentheses. Legend: ∗p < 0.1,∗∗ p < 0.05,∗∗∗ p < 0.01. We use Poisson quasi-maximum likelihood estimation because our dependent variable is an overdispersed count variable.

Table 7: Difference-in-differences estimation with modified prior deployments as the dependent variables.

	(1)	(2)	(3)	(4)	(5)
	Exploitation	Novel exploration	Near-novel exploration		
	Prior deployments	**1**[Prior deployments=0]	**1**[0<Prior deployments<Mean(Prior deployments)]		
Exposure×Post	0.13**	0.09	−0.19***	−0.01	−0.22***
	(0.06)	(0.19)	(0.07)	(0.16)	(0.08)
Model	Poisson quasi-maximum likelihood	Logistics	Logistics	Logistics	Logistics
Sample	Prior deployments > 0	All trials	Prior deployments > 0	Below-median HHI = high technological diversification	Above-median HHI = low technological diversification
N	8,961	9,723	8,961	4,477	4,484
Biologic	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y
Therapeutic area fixed effects	Y	Y	Y	Y	Y
Firm fixed effects	Y	Y	Y	Y	Y
Phase fixed effects	Y	Y	Y	Y	Y

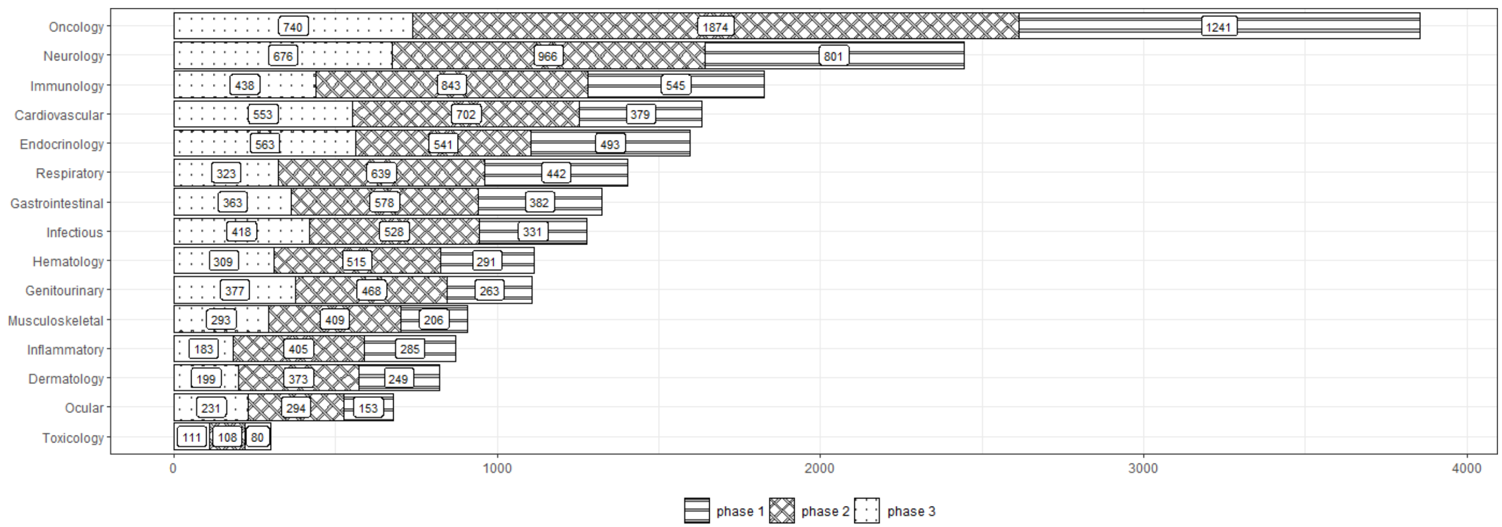
Notes. Robust standard errors (clustered at the firm level) are presented in parentheses. Legend: $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$. We use Poisson quasi-maximum likelihood estimation when the dependent variable is an overdispersed count variable. We use logistics estimation for an indicator dependent variable.

Figure 1: Clinical trials as reported in ClinicalTrials.gov and Cortellis.



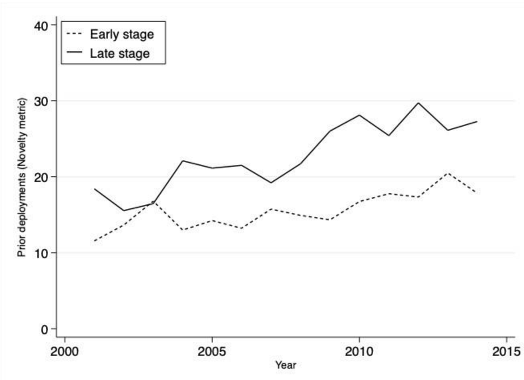
Notes. The vertical dashed lines show the date of various policies which could impact trial registrations. The solid line shows the number of clinical trials registered on Clinicaltrial.gov. The dotted line shows the number of trials in the Cortellis dataset.

Figure 2: Distribution of clinical trials by therapeutic area and phase.

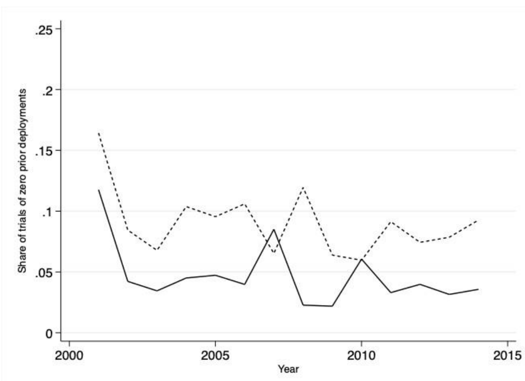


Notes. This figure shows the distribution of clinical trials by therapeutic area and phase. Each bar represents the number of trials associated with the therapeutic area. Each bar is divided into stacks representing the number of trials associated with the three phases of clinical trials.

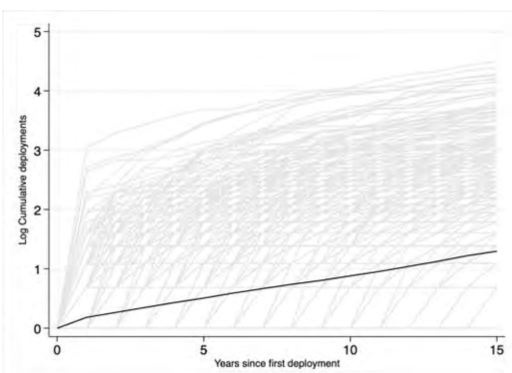
Figure 3: Patterns of technology deployment.
(A) Average prior deployments in early- and late-stage trials.



(B) Proportion of trials with purely novel technology (Prior deployments=0) in early- and late-stage trials.

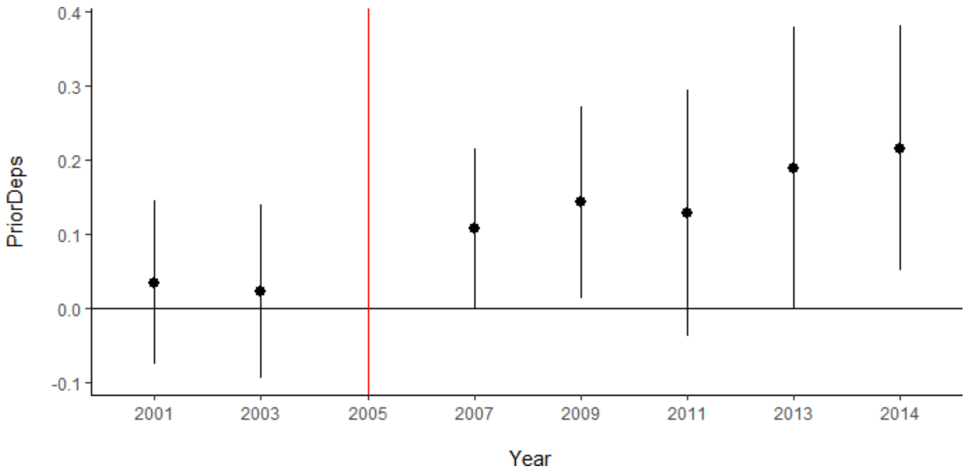


(C) Cumulative deployment since the first deployment.



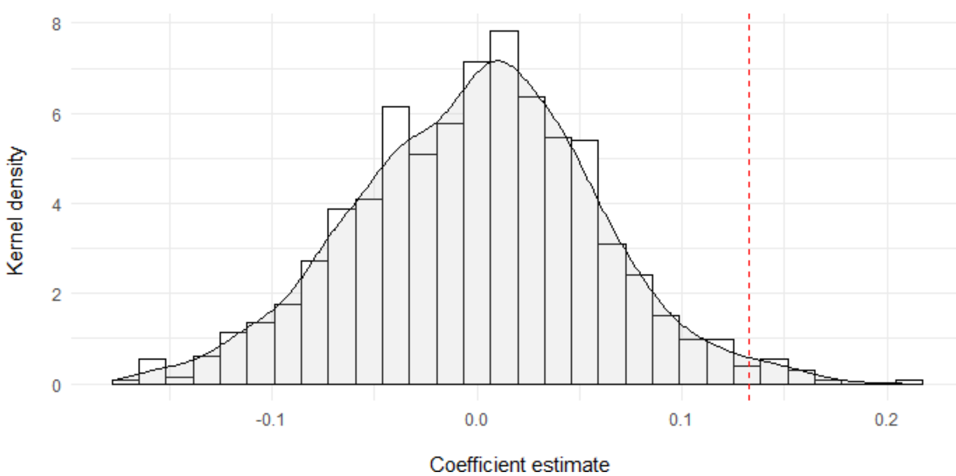
Notes. Figure A. This figure shows the average number of prior deployments for early- and late-stage trials. Late-stage trials (Phase 3) shown by the solid line exhibit larger values of prior deployments compared to early-stage trials (Phase 1 and Phase 2) shown by the dotted line indicating that the drugs entering late-stage trials are less novel than those entering early-stage trials. Figure B. This figure shows the percentage of trials that are based on a technology that has never been used before. Figure C. This figure shows each technology's log number of cumulative deployments as a function of the number of years since the first deployment. The dark solid line reflects the average across all technologies. The slope of this line indicates that each additional year since the first deployment is associated with about a 6.3% increase in the number of cumulative deployments. Alternatively, each percentage increase in the number of deployments is associated with 0.16 years (1/6.3) of technological lag.

Figure 4: Event study specification: difference-in-differences effects over time.



Notes. This figure reports the event study version of Table 4. Each marker represents the coefficient of the interaction between Exposure and an indicator variable for each two-year period, with each period ending in the shown year. For example, the marker shown for 2001 represents the estimated difference-in-differences effect for 2000 and 2001; for 2003, the difference-in-differences effect is estimated for 2002 and 2003, and so on. The interaction coefficients are calculated relative to that of the year 2005. Bars represent 95% confidence intervals.

Figure 5: Falsification test results



Notes. This figure plots the distribution of estimated coefficients across 1,000 placebo experiments. In each of these, we randomly sample Exposure from a standard normal distribution and estimate Equation 2. The dashed vertical line denotes our main differences-in-differences estimate (column 2, Table 4), which yields an implied p-value is 0.011 given the empirical distribution of falsified counterpart estimates.

DECISION SCIENCES INSTITUTE

How Firms Learn to Manage Tensions in IT Outsourcing Relationships

Elvis Ngah
Vrije Universiteit Amsterdam
Email: e.n.ngah@student.vu.nl

Brian Tjemkes
Vrije Universiteit Amsterdam
Email: b.v.tjemkes@vu.nl

Henri Dekker
Vrije Universiteit Amsterdam
Email: h.c.dekker@vu.nl

Valerie Duplat
Vrije Universiteit Amsterdam
Email: v.duplat@vu.nl

ABSTRACT

Not knowing how to manage tensions (e.g., localization-globalization) in IT outsourcing (ITO) relationships substantially damages performance and largely explains the high occurrence of contract breaches in the industry. Extant ITO research offers various tension management strategies but little on partners learning how to implement them. Through the lens of organization learning, we conducted a longitudinal case study involving a client organization that switched vendors to outsource the same IT operation. Our findings permit us to offer insights into the learning dynamics, including feedback loops, governance, separate learning paths for transient and persistent tensions, and the need to institutionalize tension management.

KEYWORDS: Relationship dynamics, IT outsourcing, (inter)organizational learning, learning patterns, tension management

INTRODUCTION

Information technology outsourcing (ITO) arrangements bring about significant tensions – elements that seem logical individually but contradictory when juxtaposed (Smith & Lewis, 2011). Partner firms face tensions such as localization vs. globalization (Pero, Rossi, & Terzi, 2018), continuation vs. switching (Peukert, 2019), and exploitation vs. exploration (Aubert, Kishore, & Iriyama, 2015). Prior ITO research highlights that an ITO relationship cannot pursue its goals and satisfy partners' interests if the partners lack sufficient knowledge of how to manage the tensions emerging in their relationship (Bapna, Gupta, Ray, & Singh, 2016; Bui, Leo, & Adelakun, 2019; Gozman & Willcocks, 2019). This inadequate knowledge is considered to largely explain the high failure rate, manifested by contract breaches and clients switching vendors (Susarla, 2012; Smite & Moe, 2020).

ITO research has provided an understanding of strategies for managing tensions in ITO relationships. These include process strategies such as interfirm learning (Mani & Barua, 2015; Kranz, 2021) and adopting ambidexterity (Cao, Mohan, Ramesh, & Sarkar, 2013; Du & Pan, 2013); and structural strategies such as multisourcing (Bahli & Rivard, 2013; Lioliou, Willcocks, & Liu, 2019) and integrating client and vendor resources (Soderberg, Krishna, & Bjorn, 2013; Kim, Anand, Larson, & Mahoney, 2019). While understanding these tension management strategies is insightful, less is known about how ITO partners learn to deploy them. Learning is critical for effectively deploying these strategies since occurring tensions are dynamic, requiring continuous response and adaptation from the partners to achieve the ITO engagement goals (Levinthal & March, 1993; Flores, Zheng, Rau, & Thomas, 2012; Ngah, Tjemkes, & Dekker, 2023).

To develop an understanding of how ITO partners can learn to manage tensions and, thus, acquire the knowledge needed to effectively deploy a tension management strategy, we rely on an organizational learning (OL) framework based on learning from experience (Argote & Miron-Spektor, 2011) and knowledge acquisition through successive learning subprocesses (Huber, 1991; Flores et al., 2012). Considering the dynamic nature of ITO relationships, Koo et al. (2017) posit that learning from experience (Argote & Miron-Spektor, 2011) is a powerful lens for examining these engagements. Since ITO relationships are variably complex (Kern & Willcocks, 2002; Bui et al., 2019; Silva, Poleto, de Gusmao, & Costa, 2020), learning experiences could be direct, indirect, a failure, or rare (Echajari & Thomas, 2015). While ITO studies mainly emphasize learning from direct and indirect experiences (Mani & Barua, 2015; Handley, Skowronski, & Thakar, 2022), studies on learning from failures (Moe, Smite, Hanssen, & Barney, 2014) and rare events are scant. Converting such experiences into knowledge involves the learning subprocesses of acquiring, distributing, interpreting, and integrating information and storing and institutionalizing knowledge through single or double feedback loops (Huber, 1991; Flores et al., 2012).

To shed light on the learning process, we address the following research question: *How do firms learn to manage tensions in client-vendor ITO relationships?* Our motivation for this question follows from the limited understanding of how partners learn to deploy a tension management strategy in ITO relationships and from the scant literature on learning from rare experiences in these variably complex engagements where failures resulting from tensions are common. To answer this question, we employ a process methodology (Langley, 1999) through an in-depth longitudinal case study (Eisenhardt & Graebner, 2007) involving a large US-based healthcare provider that switched vendors to outsource the same IT activity. Using the lens of OL (Argote & Miron-Spektor, 2011), we collected data using semi-structured interviews, allowing us to be open to any deviation from an assumed sequence of OL subprocesses (Flores et al., 2012). To accommodate this openness, we employ the Gioia methodology (Gioia, Corley, & Hamilton, 2013) to analyze the data. Our findings show learning to involve a sequence of interrelated subprocesses – *experiencing tensions*, *sharing experiences, comprehending tensions, memorizing tensions,* and *addressing tensions.* Based on these findings, we offer a model of learning dynamics characterized by feedback loops and governance, different learning paths conditional on the transient or persistent nature of occurring tensions, and institutionalization of knowledge on tension management.

We make two theoretical contributions and provide two managerial recommendations. First, we bridge the gap between known tensions in ITO relationships and known strategies to manage them. We show how ITO partners can acquire knowledge to deploy a tension management strategy through an interplay of the identified learning subprocesses. This interplay highlights that learning to manage transient tensions follows an informal path, whereas learning to manage persistent tensions follows a formal path. Second, by relying on a rare and heterogeneous (Zollo & Reuer, 2010) case, we extend research on the understudied phenomenon of ITO relationship failure and termination. Our case study provides an understanding of the termination of an existing relationship with one vendor by switching to another through learning to manage tensions. For practitioners, our findings signify the importance of accounting for tension management and the associated learning subprocesses. Our results enhance the understanding of what activities or learning subprocesses managers can implement and how to do so. We also recommend client and vendor managers to consider where and when their actions contribute to learning about effective tension management. Our findings suggest both partners should pilot the initial learning stages (i.e., *experiencing tensions*, *sharing experiences*, and *comprehending tensions*), and the client should lead in managing the later stages (i.e., *memorizing* and *addressing tensions*).

THEORETICAL BACKGROUND

Tensions in ITO Relationships

The ITO literature characterizes ITO relationships as variably complex and dynamic interorganizational relationships replete with tensions (Ngah et al., 2023). A tension represents two elements that seem logical when considered individually but opposing or incompatible when juxtaposed (Smith & Lewis, 2011). For example, in outsourcing for cost-effective innovation, extant studies have shown that these relationships may experience a tension in exploration (including value creation and flexibility) vs. exploitation (including value capture and efficiency) (Miranda & Kavan, 2005; Aubert et al., 2015). While the client explores the vendor's new ideas to ensure future growth, it also exploits the vendor's existing ideas for profitability. Exploration and exploitation are considered incompatible since they compete for the same (inter)organizational resources and require different mindsets and routines. Similarly, in efficiency-oriented offshoring for software development, a client may need a software product for its local market, while the vendor may be an organization (of a multinational firm) at a geographically distant location. This introduces a tension in localization vs globalization (Geppert, 2005; Pero et al., 2018) because such relationships may face cultural and time zone differences, hindering productivity and efficiency.

Tensions also occur in managing and governing ITO relationships. Tiwana (2013) posits that a firm faces a tension in IT vs business knowledge in terms of specializing in its own domain while maintaining knowledge in the partner's domain. As the vendor generally has more IT knowledge and the client has more business knowledge of the outsourced service, it is critical that the parties reduce the respective knowledge asymmetries to achieve better outcomes. Similarly, firms are confronted with a tension about who of (or to what extent) IT and business executives should manage their ITO engagements (Chakrabarty & Whitten, 2011). Also, governing an ITO relationship entails accommodating its dynamic nature through contractual mechanisms such as control (Kotlarsky, Oshri, Dibbern, & Mani, 2020), introducing a tension in formal vs informal control (Gopal & Gosain, 2010; Wiener, Remus, Heumann, & Maehring, 2015). Formal control mechanisms are stipulations in ITO contracts for controlling processes and expected outcomes that limit opportunism and other unwanted behavior. Whereas informal control mechanisms such as clan control and social interactions are also known to influence project performance positively, they can create an avenue for opportunism.

Recognizing the existence of tensions and their implications for ITO relationship development and performance, some studies have suggested process and structural strategies to manage them. Process-oriented studies highlight that tensions can be managed by learning from past and current relationships (Moe et al., 2014; Mani & Barua, 2015; Ali & Khan, 2016; Kranz, 2021), by ambidexterity (i.e., balancing both ends of a tension; (Cao et al., 2013; Du & Pan, 2013; Teo & Bhattacherjee, 2014; Oshri, Kotlarsky, & Gerbasi, 2015), by equivocation (i.e., adopting a wait-and-see contract renegotiation process; (Heiskanen, Newman, & Eklin, 2008), and by liminality (i.e., attaining a liminal (transitional) stance between both ends of a tension; (Nicholson, Babin, & Briggs, 2017). Structural-oriented studies, on the other hand, suggest that tensions in ITO relationships can be managed by integrating client and vendor resources (Duhamel *et al.*, 2018; Ghosh and Scott, 2009; Kim *et al.*, 2019; Soderberg *et al.*, 2013), by integrating their capabilities (Karimi-Alaghehband and Rivard, 2020; Kim and Chung, 2003; Wang and Wang, 2019), by adjusting the ITO contract to accommodate new and perceived behaviors (Akkermans, Van Oppen, Wynstra, & Voss, 2019; Kranz, 2021), and by multisourcing (i.e., sourcing the same IT activity from more than one vendor; (Bahli & Rivard, 2013; Lioliou et al., 2019).

While the literature provides an understanding of these tension management strategies, it provides less understanding of how ITO partners learn to deploy them. Learning is highly relevant as it enables decision-makers to opt for a better management strategy, allowing the partners to adapt quickly, effectively, and continuously to unexpected changes that accompany these tensions. Accordingly, we adopt an OL framework to address this gap in the literature.

Experiential Learning and ITO

OL provides a powerful lens for studying how organizations can increase the effectiveness of their actions for continuous change and adaptation in response to dynamic challenges (Levinthal & March, 1993; Flores et al., 2012). According to Beer et al. (2005), an organization can meet its strategy amid dynamic challenges if it is open to learning to manage the tensions it experiences, which otherwise could prevent it from realizing its strategy. The current study adopts an OL framework that emphasizes learning from experience, defined by Argote and Miron-Spektor (2011) as a change that occurs as an organization acquires experience. This experience could be direct, indirect, a failure, or rare (Echajari & Thomas, 2015).

Prior research has emphasized learning from direct and indirect experiences. Learning from direct experience involves learning by doing (Zollo & Winter, 2002; Annosi, Martini, Brunetta, & Marchegiani, 2020) and is salient in IORs (Tsang, 2002; Zollo & Singh, 2004; Kale & Singh, 2007), including ITO relationships (Koo et al., 2017). Research in this context argues that learning is deliberate and that experiences are dynamic (i.e., they evolve) due to the complexity of these relationships. In contrast, learning from indirect experience, also known as vicarious learning, pertains to learning from external sources, involving the transfer of knowledge developed by one entity to another. Indirect learning may, amongst others, originate internally (from different organizational units that perform comparable activities) and externally (from relationships with other organizations) (Echajari & Thomas, 2015; Argote & Hwang, 2016).

Prior research has also emphasized the relevance of learning from failure and rare experiences. Failures are common in complex environments (Edmondson, 2011), such as in ITO projects (Carmeli & Dothan, 2017), where complexity relates to the existence of tensions (Bui et al., 2019). Firms can learn more effectively from failures (than from successes) and exploit what they learn (Madsen & Desai, 2010). Yet, significant failures rarely occur; when they do occur, they constitute rare experiences (Echajari & Thomas, 2015). Firms can learn during a rare event (from direct experience) or after such an event (from indirect experience) (Lampel, Shamsie, & Shapira, 2009). Rare experiences tend to be heterogeneous as they hardly repeat in the same way, but the knowledge generated from a rare experience can be adapted and reused in the future (Zollo & Reuer, 2010), such as when a firm should engage in a new ITO relationship from a previous one.

ITO studies on experiential learning have mainly embarked on learning from direct and indirect experiences with little emphasis on learning from failure and rare experiences. Studies have, for example, used (in)direct learning to investigate sourcing decisions (Lacity & Willcocks, 1998; Whitten, Chakrabarty, & Wakefield, 2010; Handley et al., 2022), explicit articulation of requirements (Yakhlef, 2009), capability development for control design (Dekker & Van den Abbeele, 2010), knowledge management (Chua & Pan, 2008; Park, Im, & Kim, 2011; Chen, McQueen, & Sun, 2013), relationship management (Brandl, Jensen, & Lind, 2018; Mihalache & Mihalache, 2020), and relationship outcomes (Bell, Bradley, Fugate, & Hazen, 2014; Mani & Barua, 2015). The scant research on learning from failures suggests that after clients identify the reasons for failed ITO relationships, they are likely to change their sourcing strategy to insourcing and partnerships (Moe et al., 2014), providing little understanding of how learning takes place amidst complexity.

Learning from experience in complex, dynamic environments involves extracting and accumulating information to create new knowledge (Echajari & Thomas, 2015). The underlying learning activities can be organized around six interrelated subprocesses (Huber, 1991; Flores et al., 2012): *information acquisition* – sensing information from internal and external sources; *information distribution* – sharing information amongst individuals and groups; *information interpretation* – making sense from acquired information; *information integration* – forming a unified understanding from various information interpretations; *organizational memory* – embedding knowledge into a repository; and *knowledge institutionalization* – embedding learning into an organization's systems, structures, procedures, strategies, and culture. Exceptions in this sequence may occur considering the interrelations and feedback loops between the subprocesses (Flores et al., 2012). Given that ITO relationships are variably complex, dynamic (Kern & Willcocks, 2002; Bui et al., 2019; Silva et al., 2020), and susceptible to failure, we adopt these subprocesses to examine how experiences in individuals can be converted to knowledge in ITO partner firms for managing tensions occurring in their engagements.

RESEARCH METHOD

To examine how firms learn to manage tensions in ongoing client-vendor ITO relationships, we conducted a longitudinal case study, collecting information about ongoing (recent) and retrospective events. A case study is appropriate because our research question – *how* – is phenomenon-driven (Eisenhardt & Graebner, 2007). A qualitative method allows us to explore critical events, practices, and the underlying logic. Thus, our methodology aims to provide an understanding of how situations emerge and develop over time through qualitative (process) data sources (Langley, 1999). Our process research relies on organizational learning theory (OLT) as a theoretical lens to capture a sequence of events based on informants' accounts (i.e., semi-structured interviews), online sources, and observations. It is of particular importance to conduct the study as longitudinal research to unravel the underlying dynamics of the phenomenon. We employed a single case study as our research strategy to obtain an in-depth and thorough understanding of the dynamics in a single setting (Eisenhardt, 1989). To ensure quality in terms of construct validity, internal validity, external validity, and reliability (Yin, 2018), we used several tactics at different phases of the research. Table 1 summarizes these tactics.

Table 1: An Overview of Tactics Used to Ensure Quality in the Study	
QUALITY CRITERIA	TACTICS
External validity *(research design)*	- We used organizational learning theory as a lens for the study. - We chose a case study methodology because we address a "how" question. - Our rationale for the case is that it is longitudinal and rare by involving active vendor switching in sourcing IT functions.
Reliability *(data collection)*	- We developed a case study protocol to guide data collection. - We used ATLAS.ti software as our case study database for our chain of evidence, including interview transcripts and available online media articles.
Construct validity *(data collection)*	- We used multiple sources of evidence for data triangulation, including semi-structured interviews, online sources, and non-participant observation. - We had the key informant review a draft of the case study report.
Internal validity *(data analysis)*	- We derived the research framework of the study from the literature on experiential learning to guide data gathering and analysis.

Research Setting

Our longitudinal case study involves a client firm, *Alpha*, in separate ITO relationships with vendors *Beta* and *Gamma*. *Alpha* is a US-based health service provider committed to efficiency and innovation. *Beta* is an Indian multinational IT service firm offering innovative end-to-end solutions for healthcare providers. *Gamma* is a French multinational IT service and consulting firm providing unified communication, cloud, big data, cybersecurity, and transactional services.

To better serve its growing patient base, *Alpha* saw the need to develop *custom software* – a system of digital telehealth tools (website and mobile app) – to give patients access to primary and specialty care, including scheduling visits with doctors on-premise or on-video, communicating with doctors through messaging, managing insurance policies, and requesting prescriptions and refills. This system acts as an interface between *Alpha's* patients and medical staff aimed at overcoming adversities such as high service costs, long appointment waiting times, and transportation costs. It also supports expanding *Alpha's* capacity to handle patients' requests in a way that fits different patient lifestyles. By providing patients with online access to *custom software, Alpha's* goal is to improve access to high-need specialty care services, meet unique patient needs through access to resources and education that enable adapting and improving health outcomes, and potentially avoid disease-related complications.

Alpha started developing this software using in-house IT staff. After experiencing high costs of attracting new IT staff with domain knowledge, it entered a sourcing arrangement with the *Beta* in June 2016, with *Beta's* resources augmenting *Alpha's* in-house IT staff. The ITO relationship met delivery expectations for three years until *Alpha* decided to terminate it in favor of a new vendor, *Gamma*. The switching from *Beta* to *Gamma* started in November 2020 and continued until June 2021, when the *Alpha-Beta* contract ended. The new (*Alpha-Gamma*) relationship covered the IT staff augmentation needs for *custom software* development (as in the *Alpha-Beta* relationship) and the management of all of *Alpha's* IS, with almost eighty percent of *Alpha's* IT staff being transferred to *Gamma*. This new sourcing relationship thus followed a managed services model.

This case is particularly well suited to address our research question for several reasons. First, we had an opportunity for a rare research access (Eisenhardt & Graebner, 2007) and to collect data during the process of a client terminating its ITO relationship with one vendor and switching to another. Research on ITO relationship termination is scarce (Ejodame & Oshri, 2018), and studies on learning in ITO engagements (e.g., Mihalache and Mihalache 2019; Yakhlef 2009) have mainly relied on retrospective or cross-sectional data instead of data collection as the events unfold. Second, while prior ITO research has typically gathered data from clients or vendors, we collected data from staff of the client firm and from staff of the incoming and outgoing vendors. This provided a comprehensive analysis of the involved firms and permitted us to consider the IOR as our unit of analysis. Third, this case also enabled us to focus deeply on the relationships of *Alpha* with the outgoing and incoming vendors (*Beta* and *Gamma*), as *Alpha* is heavily involved in large-scale ITO projects using a single vendor strategy. This makes the emergence of tensions and learning to manage them well imperative.

Data Collection

Based on the OL lens used in the study, we account for learning from experience and the subprocesses through which this learning could take place. Guided by this framework, we adopted a broad approach to data collection that covered the learning experience and accommodated the emergence of new insights and potential deviations. We developed a case study protocol to guide data collection (Yin, 2018), which included an introduction to the case, the guiding theoretical framework, likely data sources (documents, records, non-participant observation, and semi-structured interviews), the interview strategy, process, and questions, and a strategy for obtaining other data. We relied on semi-structured interviews with key participants on real-time and retrospective events as the primary data source, complemented by non-participant observation and documents, to achieve data triangulation and improve the reliability of our analysis (Gioia et al., 2013). Due to the sensitive nature of the case, we were not able to access the companies' records. Instead, we relied on the companies' websites for relevant documents and insights (Vicente, Ferasso, & May, 2018).

We gained access to the case through our key informant, a Process Manager at *Alpha*, with whom we conducted an initial interview to establish whether the case would suit the phenomenon under investigation. This interview revealed the *Alpha-Beta* relationship had existed for over three years until *Alpha* decided to switch vendors. The key informant's responsibility was to stabilize the outsourcing project, especially during the switching period. The interview insights suggested that the emergence of various tensions and learning how to manage them would be important topics in this case. Eventually, the key informant helped identify other informants from both *Alpha* and *Beta* for our first round of interviews, in which we captured retrospective data about the development of the *Alpha-Beta* relationship. We also actively observed the vendor-switching phase, during which we conducted a second round of interviews, capturing insights as events unfolded. As we progressed in the interviews, we adapted our interview questions based on various evolving insights and for clarification, also leading to interviewing new informants. This process continued until we reached saturation, with additional interviews providing little new information.

The informants' roles ranged from lower to senior management, including Technical Lead, Team Lead, Relationship Manager, Process Manager, Senior Manager, and Vice President (see Table 2). The same client informants were involved in both relationships, while three vendor informants (marked by *) were transferred from the old vendor *(Beta)* to the new one *(Gamma)*. This distribution permitted us to collect data from client and vendor informants who participated in all the stages of the ITO engagement relevant to this study. Due to the strategic, sensitive, and complex nature of the information requested, we sent an outline of the case protocol to informants before the first interviews to permit them to prepare and revisit archives to support their statements (Voss, Tsikriktsis, & Frohlich, 2002). Besides the interviews, we followed up so informants could verify emails and internal documentation to retrieve information relevant to our analysis. In addition, we complemented the interviews with online news articles and our observations of the evolution of the case.

Table 2: List of Informants

PERSPECTIVE	INFORMANTS	NR OF INTERVIEWS
Client	Process Manager (key informant)	6
	Vice President	2
	Senior Manager	2
	Team Lead	2
Vendors	Relationship Manager*	2
	Relationship Manager*	2
	Technical Lead*	2
Total		18

The entire data collection took place from January 2021 to April 2022 and comprised eighteen recorded conversations, eleven online sources (vendor website and news articles), and observations by the lead author. Interviews were recorded via Zoom meetings, each lasting between 30 and 60 minutes. We made notes to record the impressions and comments from informants during the interviews. After analyzing the interview transcripts, we conducted multiple follow-up interviews with the same informants or with new leads to clarify and hone details or questions arising from our earlier interviews. We also followed up to clarify some details from the online sources.

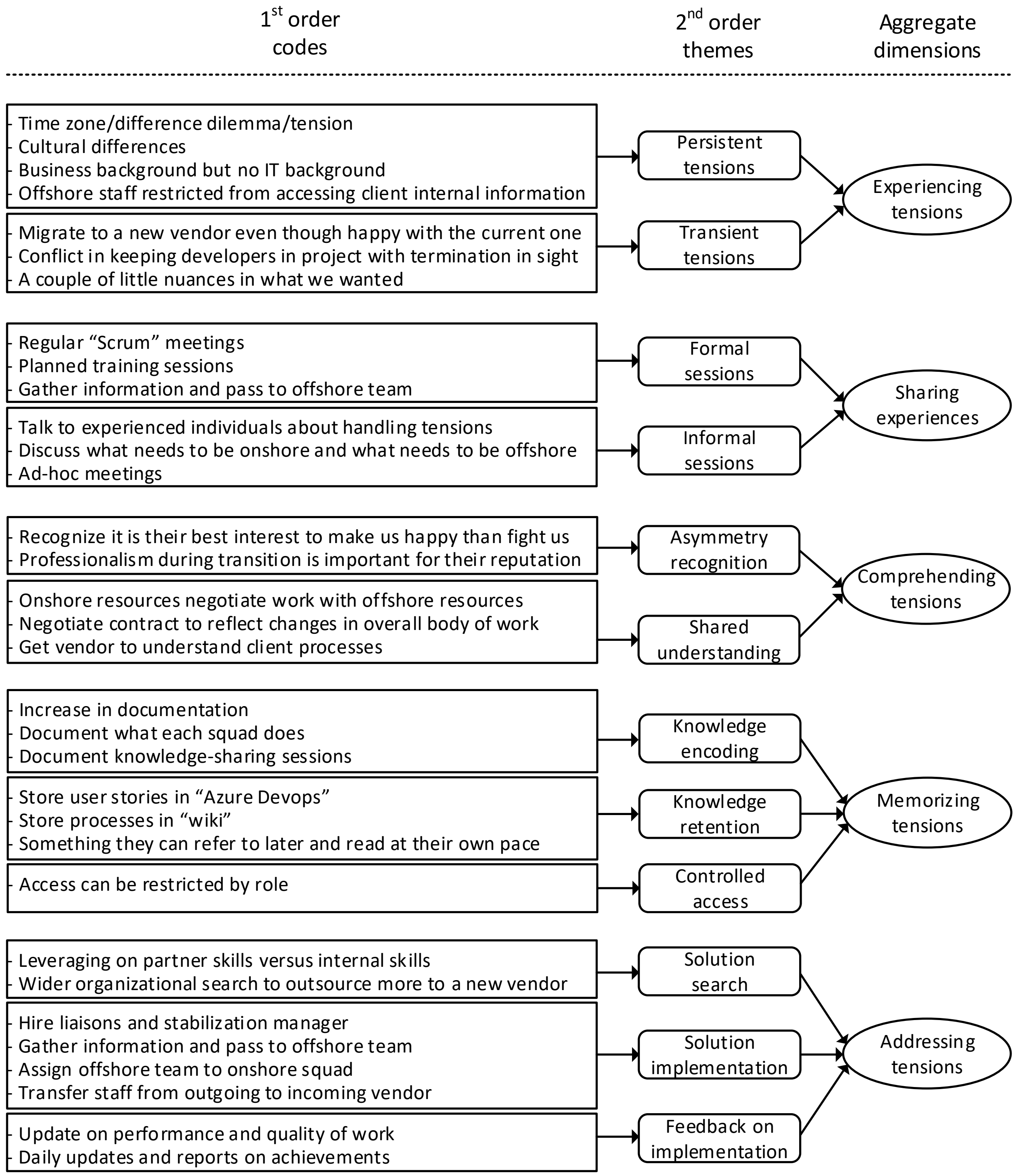
Data Analysis

As a first step in the analysis, we sought to establish if the case was suited for the study and also to familiarize ourselves with the case. The first author analyzed the first two interviews with the key informant. This was also followed by an analysis of the data we collected from online sources about the case. This initial analysis helped us restructure the case protocol for our first round of interviews with informants, which we analyzed to obtain our first set of codes. These codes (in the informants' own words) and observations of how the case evolved were discussed with the key informant, during which we noted down ideas and updated the case study protocol for the next round of interviews.

After completing data collection and being open to new insights to accommodate potential deviations from the OL framework, we used an inductive approach following the Gioia methodology to analyze the data (Gioia et al., 2013). We progressed from the empirical details in the informants' own words and online sources toward a data structure that could be compared, categorized, discussed, and analyzed through our chosen lens (Langley, 1999), OLT. To enhance qualitative rigor, we analyzed and organized the data into first and second-order categories to facilitate their aggregation into a more structured form.

Figure 1 shows the outcome of our data structuration process. For the first-order analysis, we used ATLAS.ti software, and open-coded the data by adhering faithfully to in vivo or informants' terms. Next, we sought similarities, differences, and linkages among the first-order codes to develop second-order themes. For the second-order analysis, we conducted axial coding by reducing the data into a theoretical level of themes guided by the OL lens. The analysis becomes more explicitly theory-driven in this stage, focusing on nascent concepts that leap out of the empirical data. Once we obtained a workable set of concepts for the second-order themes, we conducted theoretical coding by assessing the semantic relationships of the second-order themes and aggregating them into relevant learning or newly identified subprocesses. This process involved reviewing and discussing the classifications, relations, aggregations, and labels for acceptance by all authors.

Figure 1: Data structure



RESULTS

This study adopts the OL framework of learning from experience (Argote & Miron-Spektor, 2011; Echajari & Thomas, 2015) to investigate how an ITO client and its vendor(s) learn to manage tensions in their relationship. We identified five dimensions relevant to this learning in this case. These include the subprocesses of *experiencing tensions, sharing experiences, comprehending tensions, memorizing tensions,* and *addressing tensions*, some of which align with (pertain to) the conceptualized OL subprocesses to acquire knowledge (Flores et al., 2012; Huber, 1991) and some not. We sequentially report the dimensions we identified (i.e., the learning subprocess to manage tensions) and then summarize an overview of their alignment with the conceptualized OL subprocesses.

Experiencing Tensions

Our findings reveal that the learning process starts with ITO partners experiencing duality – when information is viewed from two perspectives that logically oppose each other, i.e., are in tension. In this regard, we found that partners experienced *persistent* and *transient tensions*.

*Persistent tensions***:** *Persistent tension* pertains to a tension that was experienced and lasted throughout the outsourcing relationship. For example, after *Alpha* signed the outsourcing deal with the vendor Beta in June 2016, *Alpha* and *Beta* staff started experiencing a time zone duality when performance concerns emerged, which persisted throughout the relationship. This duality can be explained by the fact that the client, *Alpha*, based in the USA, needed the developed software for local deployment, whereas the Indian-based vendor, *Beta*, has a global presence, with most of its staff in India and a few in the USA and elsewhere worldwide. Upon experiencing this *localization-globalization* tension in terms of longer delivery times, a vendor Technical Lead who was onshore (in the USA) and a member of the client's software testing team noted:

> *"I would mostly depend on the offshore team to fix tiny bugs because it would be time-consuming for me. So, kind of [a] tension [because] I had to wait until I get to the offshore team in the evening so that I can explain it to them, and they can fix it immediately."*

Transient tensions: *Transient tension* pertains to a tension that occurred for a given time and was resolved during the ITO engagement. For example, in 2020, the broader client firm asked the IT organization to switch the outsourced activity (*custom software* development) from the vendor *Beta* to a new vendor, *Gamma*. However, *Alpha's* IT organization did not want to do so because it was happy with the performance of the incumbent vendor, *Beta.* This reluctance during the *Alpha-Beta* relationship is evidence of the tension, *continuation-switching*, which only lasted for a while, i.e., it was transient. Experiencing this tension is best captured by the words of the VP of the client's IT organization:

> *"The driving factor for termination was not that we ultimately got to distrust each other or were unhappy with the work. It was more of a business decision of only wanting a single vendor because we could have stuck with Beta and just brought in Gamma to do everything else, which still would have been a very big contract … But I think for economies of scale and pricing, and for a single throat to choke or a single vendor to do it all, the decision was made to migrate, even from a vendor [Beta] that we were happy with, to Gamma."*

Sharing Experiences

We found that staff shared how they experienced tensions during *formal* and *informal sessions*. During these sessions, they explained to each other the tensions they experienced and their views on the strategies to manage them.

Formal sessions: Regular meetings with structured agendas were planned to improve productivity. These formal sessions were mandatory for onshore resources and were designed to let teams plan their work, identify any obstacles at various stages of the ITO engagement, and suggest solutions to those obstacles. One such meeting is *Scrum,* during which members gave updates on project progress, achievements, and challenges. They further shared what they learned from these sessions with other team members. For example, after experiencing a time zone duality, a vendor relationship manager recounted as follows:

> *"And with this onshore-offsite model, not everyone will have all these meetings. But one person, whoever is on-site, will join all these meetings and discussions with the developers carrying all the details, and they will share it with offshore people … Everyone on-site needs to join these scrum calls and understand what's going on. You don't want to say time is wasted when you join these meetings."*

Informal sessions: Not all exchanges took place through formal meetings; several informal sessions also occurred among individuals. As challenges arose, these sessions took place on an ad-hoc basis to get to the bottom of things. During discussions, partner resources shared their understanding of the tensions they experienced and sought to understand how to manage them. The person experiencing the tension would usually initiate such meetings to get a solution from someone who might have managed such tensions in a previous experience. This is captured in another vendor relationship manager's own words:

> *"The learning kind of experience is that initially, when I experienced tension[s], I just needed to talk about it to an experienced person, with other people, [for] how to manage it."*

Comprehending Tensions

We found in this ITO engagement that comprehending the shared experiences on tensions, and the strategies to manage them is another subprocess towards learning how to manage tensions. This comprehension includes *asymmetry recognition* and *shared understanding*.

Asymmetry recognition***:*** Sometimes, there were different client and vendor dispositions from the shared experiences, which, if not handled adequately, would result in a divergence in client and vendor behaviors toward managing tensions. This prompted the partners to recognize such asymmetry, make sense of their dispositions, and interpret the tensions and tension management strategies based on the realities in the ITO engagement. For example, in 2020, *Alpha* informed *Beta* that it would terminate the ITO engagement and switch to another vendor *(Gamma)*; at the same time, it would require some onshore resources (knowledge holders) from *Beta* to transfer to *Gamma.* With termination in sight, *Beta* had a different disposition toward the ITO engagement and was unwilling to let its staff join the new vendor. In recognizing this asymmetric dispositions of *Alpha* and *Beta*, both parties made sense of the situation at hand and came to a consensus, as the VP of the client's *(Alpha's)* IT organization stated:

> *"We ultimately got them to recognize that it was probably in their best interest to let some of these people transition to us than to fight for them and make [us] an unhappy customer because if they would have fought us too much, then they would have left us as an un-referenceable customer."*

Shared understanding: The data also points towards realizing a shared understanding between client and vendor teams about their experiences with tensions and tension management strategies. As the client/vendor and offshore/onshore teams interacted, they negotiated activities and shared practices, knowledge of processes, and the use of systems and technologies. These activities helped them to incorporate knowledge about tensions and tension management that were not common to everyone. For example, a vendor relationship manager explained how assigning offshore teams to onshore squads facilitated a shared understanding of squad processes, policies, and communication technologies:

> *"We needed to get them [the offshore team] up to speed to understand all the processes in place. Each team and each squad is different in a project … and each has its own set of rules, policies, or workflows. So, based on the squad that the offshore team is assigned to, the team needs to be informed of the squad's processes and workflows.*

Memorizing Tensions

Memorizing tensions pertains to embedding knowledge on tensions in a repository so that it persists over time and is available whenever needed. We found that this subprocess entails *knowledge encoding*, *knowledge retention,* and *controlled access.*

Knowledge encoding: At the beginning of the ITO engagement, no formal rule existed to encode knowledge acquired about tensions. However, over time, it became a common practice to do so for the client and vendor resources to use the encoded knowledge whenever needed. This activity, which entails converting personnel experiences to digital documentation such as filming training sessions and creating online *Wikis*, was a first step towards ensuring that acquired knowledge could easily flow between teams and staff. As such, it, for example, contributed to managing the *localization-globalization* tension encountered by offshore and onshore teams, which were based in geographically distant regions (India and the USA) with regular personnel turnover. In this regard, the client's IT Operations Team Lead highlighted that:

> *"Our DevOps tool has the ability to create Wikis. So we have been doing that …. I've created two or three of them myself …. I feel like it is an ongoing challenge, especially with teams nowadays that have so much turnover, that creating documentation for expectations and for the processes is vitally important because the offshore team and the onshore teams tend to be having turnovers."*

Knowledge retention: The client stored encoded knowledge on tensions to ensure it was readily available when needed. We observed that the client had two main repositories for this – *Azure DevOps* and *Wiki. Azure DevOps,* a Microsoft tool, was used to store technical knowledge on project processes, which facilitated managing tensions between client and vendor teams. *Wiki*, an online organizational knowledge repository, was used to store non-technical processes on tension and tension management, among others. Considering these, retention occurs when knowledge of tensions can be stored in a repository so that client and vendor staff can readily recall and use the knowledge when needed. The client's IT Operations Team Lead emphasized the importance of this knowledge retention as follows:

> *"We did have some sort of knowledge transfer meetings, but the Wiki puts it in writing, and they can refer to it anytime they want. And it was helpful for offshore as well … I think having it written down, something that they can refer back to later and read at their own pace, helped those guys a lot."*

Controlled access: Retained knowledge was not open access for everyone in the client and vendor organizations but for designated personnel participating in the ITO engagement. The client ensured that these individuals could do so hassle-free, as further recounted by the client's IT Operations Team Lead

> *"They haven't had any trouble getting to it. I know that it can be restricted by role but I don't know anybody that's had any trouble getting to the Wikis I have created."*

Addressing Tensions

Upon learning, after mutually interpreting the tensions experienced and/or keeping a tension memory for the ITO engagement, actions to address the tensions experienced were observed. These relate to *solution search*, *solution implementation*, and *implementation feedback.*

<u>*Solution search:*</u> The first step of managing a tension was to search for its solution locally, i.e., from the current relationship where the tension occurred. If this local search did not identify acceptable solutions, then the search was expanded. For example, when *Alpha* wanted to outsource additional IT functions, such as *billing*, and noticed that *Beta* could not handle them, a search for a solution outside of the *Alpha-Beta* engagement was initiated. In search for a solution to manage this *exploitation-exploration*, *Alpha's* Process Manager mentioned:

> *"There was a larger organizational search going on into the outsourcing needs of Alpha, and that was "custom development" plus other things like back office administration and what we call revenue cycle […] Several companies were evaluated, but one came to kind of a paramount in the minds of leaders and that was Gamma because they had a broad range of skills, both accounting and customer support [...] It was also discovered that they could provide the custom development support we also needed […] So a decision was made above the Custom Development Department and then pushed down that we would be switching from Beta to Gamma.*

<u>*Solution implementation:*</u> The next step in tension management was to implement a solution, i.e., implement a tension management strategy, depending on the tension encountered. For example, to manage the *exploitation-exploration* tension, one such solution was training partner staff with the help of the established *tension memory*, as further indicated by the Process Manager at *Alpha*:

> *"There was the agreement with Gamma that they would have a broader set of responsibilities than Beta did. Now, I'll give you an example. In the area of QA [quality analysis], we had the Alpha employees doing many of these roles. And so, one of the transitions was … we used to do a lot of QA testing onshore because we thought a specific domain knowledge was needed. And what we found was that with a little bit of documentation and cross-training, we can move much more of that offshore. And so, we have offshored some things more than we thought. And that was part of the learning through the last several months as we have built this relationship with Gamma."*

<u>*Implementation feedback:*</u> After a solution was implemented, the client checked if the solution worked appropriately or if improvements were needed. This was reflected in the performance of the ITO engagement. As such, the client resorted to close control by regularly obtaining feedback on project activities. This was recounted by the Tech Team Lead of *Gamma*:

> *"Yeah, obviously, it was the client; their priority is to get the features and the tasks done in a timely manner. So, I provided them the up-to-date status and progress every day ... in the morning, we had to present our status in the squad meetings daily."*

Summary of Findings

Our case study findings indicate that to learn to manage tensions in ITO relationships, partner firm staff need to experience the tensions, share their experiences, and incorporate, memorize, and manage the tensions. These subprocesses form the basis of the learning process. While some of these subprocesses align with (pertain to) some OLT subprocesses to acquire knowledge (Huber, 1991; Flores et al., 2012), others show some misalignment or deviations (see Table 3). *Experiencing tensions* pertains to information acquisition; *sharing experiences* pertains to information distribution; *comprehending tensions* pertains to *information interpretation* and *integration*; and *memorizing tensions* pertains to *organizational memory*. Meanwhile, we could not identify any subprocess pertaining to the OLT subprocess of knowledge institutionalization. Conversely, we identified *addressing tensions* as a learning subprocess that does not pertain to any OLT subprocess. The interplay between the identified learning subprocesses and the deviations from the conceptual OLT subprocesses form an integral part of our discussion.

Table 3: Mapping of Subprocesses from Our Findings with OLT Subprocesses

IDENTIFIED LEARNING SUBPROCESS TO MANAGE TENSIONS	PERTAINING OLT SUBPROCESS TO ACQUIRE KNOWLEDGE
Experiencing tensions	Information acquisition
Sharing experiences	Information distribution
Comprehending tensions	Information interpretation
	Information integration
Memorizing tensions	Organizational memory
	Knowledge institutionalization
Addressing tensions	

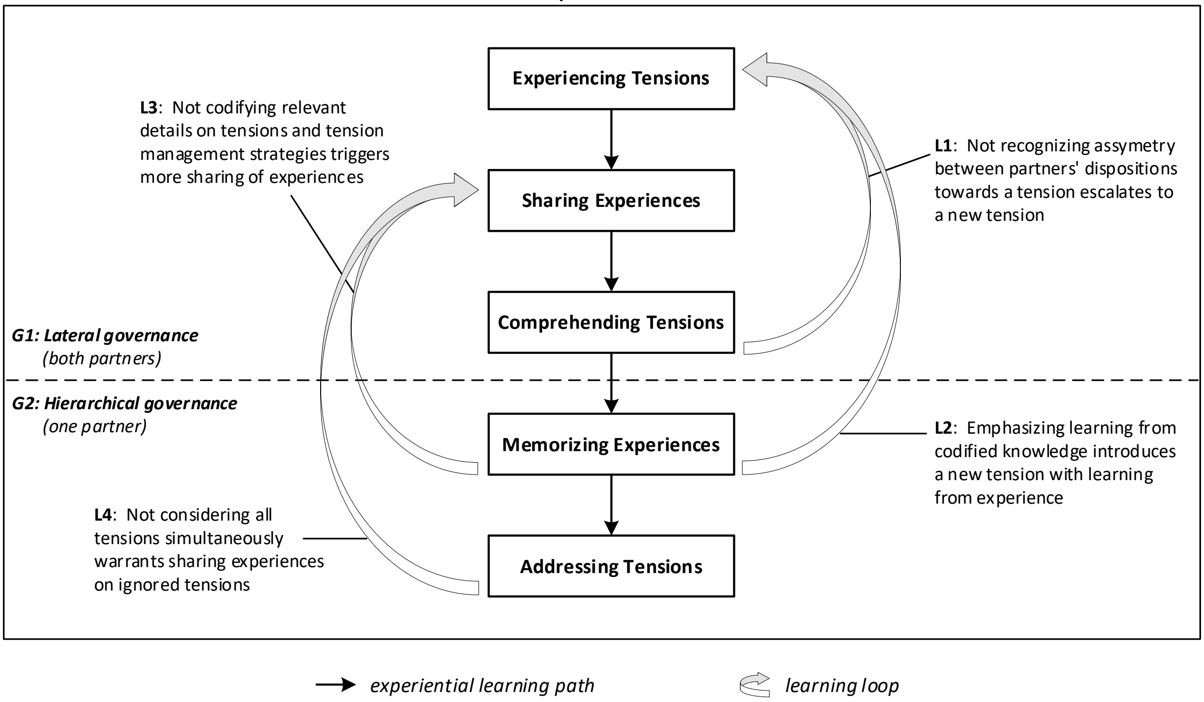
DISCUSSION

So far, we have presented the results following a linear sequence guided by the OL framework of learning subprocesses. Here, we extend our findings based on observed deviations from this sequence and insights from existing literature to elucidate the dynamics of the learning process. These dynamics entail feedback loops and governance, differential learning paths, and the need to institutionalize tension management knowledge.

Feedback Loops and Governance

The learning subprocesses observed in the case study are interrelated through feedback loops, causing the learning process to deviate from the linear sequence presented in the results section. The loops suggest that the learning is non-linear but continuous, portraying feedback into earlier subprocesses. Figure 2 visualizes these dynamics, with L1 to L4 representing identified loops and G1 and G2 representing the modes in which the subprocesses are governed. Governance mode refers to how the activities constituting the learning subprocesses are organized and coordinated, including interorganizational structures, allocation of decision-making authority, and procedures (Rodríguez, Langley, Béland, & Denis, 2007; Markus & Jacobson, 2015). These include the interorganizational actors and units performing and coordinating learning activities and review and approval processes. We detail these loops and governance modes below.

Figure 2: Interplay between learning subprocesses



L1: During the subprocess of *comprehending tensions*, the partners make sense of (and integrate) the experiences they share about tensions and tension management strategies. The activities aim to minimize asymmetry in the partners' dispositions toward tensions. However, if they cannot recognize and minimize this asymmetry, they are bound to be equivocal and have a conflict in their understanding (Gupta, Smith, & Shalley, 2006; Flores et al., 2012). For example, In 2020, *Alpha* changed its focus from exploitation (efficiency gains) to pursuing exploitation and exploration by outsourcing more IT functions (in addition to *custom software* development) and therefore took an ambidextrous disposition towards tension interpretation. However, the perception of the vendor *Beta* remained unchanged with exploitation, i.e., inert, predisposing information about exploration. This situation reflects the emergence of another tension *inertia-adaptability* – a new experience, thus a loop (L1) back to the initial subprocess of *experiencing tensions.*

L2: Learning to manage tensions in the case was achieved through experiences. During the subprocess of *memorizing tensions*, tacit knowledge within individuals was converted and accumulated to explicit (codified) knowledge. The value of this codified knowledge is achieved when individuals retrieve and use it, and by doing so, the combined learning from experience and from this codified knowledge can improve tension management. This may motivate firms to stress learning from codified knowledge and learning from experience. While these two learning modes may be beneficial for knowing how to manage tensions, their co-existence also generates a tension (Lundvall, 2016). This tension may be seen in the need to reconcile knowledge management strategies prescribing the use of IS (such as online *Wiki*) as tools for codifying and sharing knowledge on tension management with strategies mobilizing individuals for (in)formal meetings (such as *Scrum*) to share tacit knowledge about experiencing and managing tensions. The emergence of this learning tension suggests a loop (L2) back to *experiencing tensions.*

L3: The case study findings suggest that although vendor staff contributed to other learning subprocesses, they essentially did not document their tension management experiences. This led to incomplete memorization of knowledge about managing tensions and, eventually, improper tension management. This incompleteness in *memorizing tension* has been hinted at by prior research on learning (Lundvall, 2016) that some individuals may withhold/forget some information, others might not document interpreted experiences, and even those that do so might only document it partially. For effective tension management, partners need to consider learning to manage a tension in its entirety by reinforcing the sharing of any (un)intentionally withheld/forgotten information about the tension so that the knowledge can be augmented through the subsequent learning subprocesses. This triggers a loop (L3) back to *sharing experiences*.

L4: Sometimes, the partners fully learned about an existing tension but did not take an inclusive approach toward all other tensions they experienced. For example, *Alpha* and *Beta* focused mainly on managing the *localization-globalization* tension but paid little attention to addressing the emergent *exploitation-exploration* tension. Prior research on tension management (Sheep, Fairhurst, & Khazanchi, 2017; Jarzabkowski, Bednarek, Chalkias, & Cacciatori, 2022) has highlighted that relational stability is obtained when co-occurrent tensions are addressed simultaneously. Failure to do so (i.e., incompletely *addressing tensions*) results in instability (Ngah et al., 2023), which could lead to relationship failure if not handled correctly (as we saw in the *Alpha-Beta* relationship). Therefore, firms need to consider and learn how to manage all aspects of co-occurring tensions. Neglecting any aspect, such as improperly learning how to manage them or ignoring one tension, creates room for relationship instability, warranting firms to consider resharing neglected experiences about tensions. This indicates a feedback loop (L4) from *addressing tensions* to *sharing experiences.*

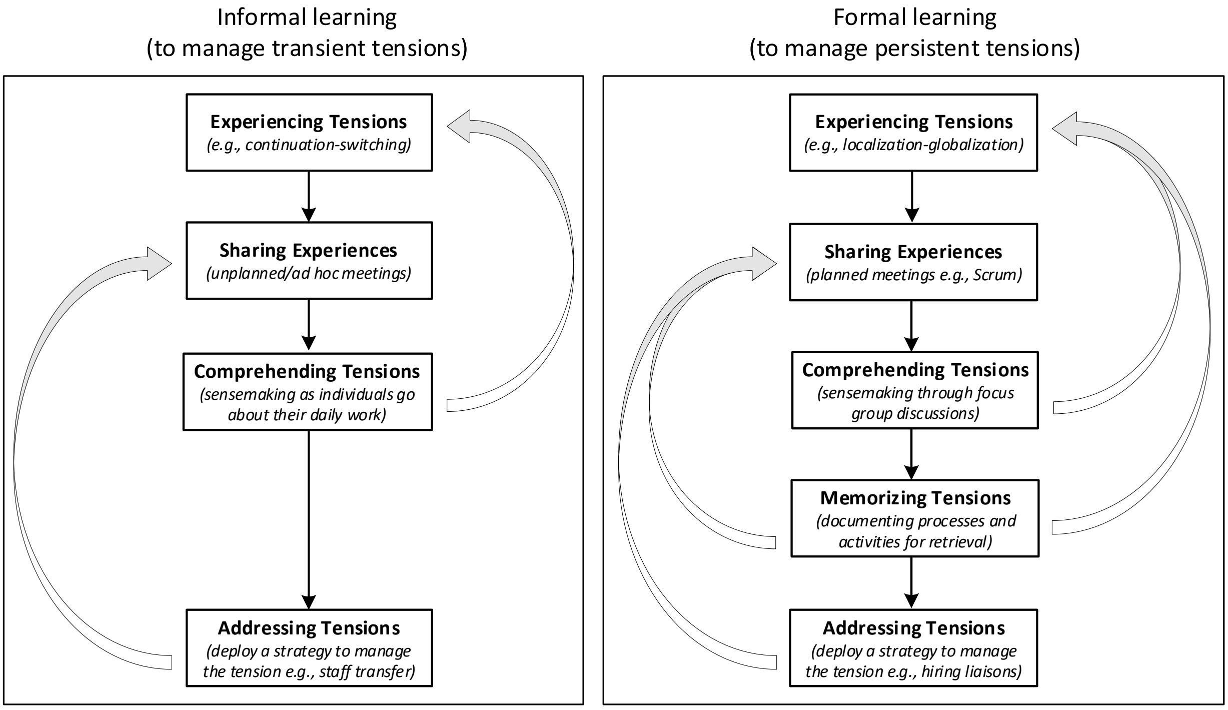
G1: Lateral governance (i.e., governance by both partners) involved client and vendor combining their efforts to share, discuss, and interpret details about the tensions they experienced, including the strategies to manage the tensions, following commonly established processes and norms. Thus, the fundamental mechanisms associated with developing a shared understanding involved client-vendor interactions (such as information exchange on tensions) and joint activities (such as training sessions), resulting in mutual learning about managing tensions as they occurred under the leadership of both partners. Thus, we posit that implementing learning processes for tension management involves lateral governance between the client and vendor parties.

G2: Hierarchical governance involves governance by one partner. As ITO relationships are dynamic (Heiskanen et al., 2008), implementing learning subprocesses requires augmentation and reinforcement to avoid becoming obsolete over time. In the case study, this was ensured by *Alpha* unilaterally establishing rules for making learning more efficient and sustainable. For example, it ensured that its staff documented processes for retention and retrieval, set controlled access to such information, and mandated vendor staff to provide regular performance feedback. Thus, augmenting and making the learning processes for tension management more efficient and sustainable followed a hierarchical form of governance (Miranda & Kavan, 2005; Rodríguez et al., 2007), where the power lies with one partner – in our case study, the client.

Informal and Formal Learning Paths

In the results, we identified different learning paths toward knowing how to manage tensions in ITO engagements. These are informal learning paths (for transient tensions) and formal learning paths (for persistent tensions), as shown in Figure 3.

Figure 3: Informal and formal learning paths



Learning to manage transient tensions follows an informal learning path. When partners experienced transient tensions, such as *continuation-switching*, learning was more about managing short-term (transient) asymmetries. Thus, there were mild traces of planned meetings but rather conspicuous unplanned and sometimes heated discussions. Learning was essentially about minimizing asymmetry in the partners' dispositions toward the transient tension for a shared understanding. This sensemaking took place through negotiations whereby as individuals of either party went about their daily work, they reflected on their dispositions and those of the other parties towards an occurring tension vis-à-vis the environmental reality. These activities corroborate the findings of extant research on informal learning (Leitch & Day, 2000; Kyndt & Baert, 2013), whereby individuals collect information and consider different viewpoints around them as they go about their activities.

Learning to manage persistent tensions follows a formal learning path. For example, *localization-globalization* was a persistent tension experienced in the *Alpha-Beta* and the *Alpha-Gamma* relationships. In either case, sharing experiences occurred through planned and regular meetings, including weekly *Scrum*. Thus, *comprehending tensions* was through focus group discussions, and this was accompanied by frequent documentation of processes, meetings, and knowledge for storage and retrieval towards dealing with tensions, i.e., *memorizing tensions*. This is consistent with prior scholarly work on learning (Mani & Barua, 2015; Lundvall, 2016), stipulating that these activities are formal, usually requiring formal expert knowledge and formal IS infrastructure (such as *Scrum Master* and *Wiki*, respectively, in the case study).

While we have elaborated that these learning paths are independent and conditional on the type of tension (transient or persistent), they do not exist in isolation. Given that transient tensions can exist alongside persistent tensions, both paths should be considered together. This aligns with the literature on interorganizational paradoxes (e.g., Putnam *et al.*, 2016), which posits that firms must take an inclusive approach regarding all the tensions they face. Consequently, this introduces a new tension in *informal-formal learning*. In this regard, we posit that effectively managing tensions in ITO relationships warrants a reconciliation between the two forms of learning.

Need to Institutionalize Learning to Manage Tensions

Our results suggest that *addressing tensions* is a learning subprocess (rather than a consequence of learning). This is because it has some implications for learning by containing elements such as solution search and implementation feedback (see Figure 1). In addition, it has a loop (L4) that feeds back to *sharing experiences* (see Figure 2). This extends prior understanding of organizational learning subprocesses (Huber, 1991; Flores et al., 2012), which may conceive *addressing tensions* as an outcome of learning. It also corroborates the literature on interactive control by manifesting as an input to *sharing experiences* (see Figure 2) for continuous learning. This literature (e.g., Simons, 1994; Simons and Dávila, 2021) suggests that managers use performance information to learn about (relational) activities that provide feedback into earlier stages from which they can make strategic and operational adjustments. Our results reveal that when partners experience tensions and, over time, build the knowledge to manage them, this knowledge feeds back into earlier stages, warranting the partners to adjust how they (re)share experiences on tensions. Thus, learning to manage tensions is an evolutionary process that should culminate in routinized actions to institutionalize the learning while addressing new tensions (Visnjic, Jovanovic, & Raisch, 2022).

However, our results did not identify any subprocess that aligns with the OLT subprocess of knowledge institutionalization (see Table 3), the subprocess that ensures that routinized actions occur by embedding what has been learned by individuals and groups into the organizations' systems, structures, procedures, strategies, and cultures (Crossan, Lane, & White, 1999; Flores et al., 2012). This may be explained by the fact that learning to manage tensions in our case study was efficiency-driven (i.e., to avoid repetitions), with no foreseen benefits for making the learning a routine. It could also be that the partners were short-sighted and did not make effective use of the acquired knowledge on tension management by not institutionalizing the learning (Teo & Bhattacherjee, 2014). However, the existence of the subprocess of *memorizing tensions* for learning to manage persistent tensions (see Figure 3) indicates that the learning was formalized in systems, though not in how the systems were used, such as making the learning (from the codified knowledge) a routine. Although institutionalizing the learning did not occur, we infer that an antecedent existed for institutionalizing the acquired knowledge about managing tensions.

While knowledge institutionalization might not be particularly relevant in our case, it may be important in other ITO settings (Teo & Bhattacherjee, 2014). Knowledge institutionalization facilitates interorganizational learning, for example, about managing tensions, when partners intertwine their activities, i.e., cooperate, toward the desired goal (Holmqvist, 2004; Jones & Macpherson, 2006). It also guides partners on how to use (exploit) the acquired knowledge and gain (explore) new insights for updating (renewing) existing knowledge (Lengnick-Hall & Inocencio-Gray, 2013). As such, it exists at the organizational level, warranting partner firms to have appropriate systems, routines, and procedures. For example, our study results identify *Wiki* as a system a partner organization should have in place, which aligns with prior IS research on organizational knowledge institutionalization (Bibbo, Michelich, Sprehe, & Lee, 2012; Argyris & Ransbotham, 2016). In addition, our study results highlight a hierarchical form of governance at the later stages of learning, suggesting that the governing partner must ensure proper procedures are in place. This may include understanding the organizational culture and routines for accessing, writing, updating, and easily using the knowledge base under project leadership (Argyris & Ransbotham, 2016).

CONCLUSIONS AND IMPLICATIONS

In this study, we aimed to understand how client and vendor firms learn to manage tensions in their ITO engagement. We have used a case involving the sourcing of IT functions by a US-based ITO client (*Alpha*) from an India-based vendor (*Beta*), which later switched to a French-based vendor (*Gamma*). In this case study, we have used the OL frameworks of learning from experience and learning subprocesses to identify the subprocesses for learning to manage tensions that can be experienced in an ITO engagement. Based on our research findings, we have discussed a model that provides insights into the learning dynamics. This model provides a holistic view based on the identified learning subprocesses and their interplay and has research and managerial implications, which can be used as points of departure for future research.

Theoretical Contributions

We make two theoretical contributions. First, while prior ITO research has provided an understanding of tension management strategies in client-vendor relationships but little about learning to deploy them, our study contributes a theoretical grounding by providing insights into the subprocesses through which firms can learn to manage tensions in ITO relationships. These learning subprocesses and the interplay between them generate the knowledge to deploy a tension management strategy. There is a different learning path for transient tensions and for persistent tensions, i.e., this knowledge is generated differently for different types of tensions. As ITO relationships are dynamic, this knowledge should contribute to the firms' competitive advantage by helping them dynamically manage these tensions. Future research could, therefore, investigate how firms can build on this knowledge to develop a dynamic capability for tension management in ITO relationships.

Second, relationship failure and termination in ITO are understudied, and organizational learning from failure amidst heterogeneity is rare. By focusing on a rare and heterogeneous setting, our unique and longitudinal case study has provided a new contextual factor for terminating an existing ITO relationship through learning to manage tensions while switching vendors. As such, our study adds to the few ITO studies on relationship failure and termination (e.g., Carmeli and Dothan, 2017; Koo *et al.*, 2017). The case is heterogeneous because the *Alpha*-*Beta* relationship is a resource augmentation engagement, whereas the subsequent *Alpha*-*Gamma* relationship is a managed services engagement. Future research could, for example, use the findings of this study as a basis to investigate how firms can use the learning experience from a previous relationship and during the vendor switching phase to effectively manage tensions (i.e., deploy a tension management capability) in a new relationship.

Managerial Recommendations

Our study also has two practical implications for managers of ITO engagements. First, managers need to account for tension management, what learning subprocesses are associated with it, and how to deploy them. The study suggests that knowing how to manage tensions is a prerequisite for tension management in ITO relationships. Instead of directly rushing to manage a tension (or contradiction) when it occurs, practitioners should first take a strategic view to understand what type of tension is experienced and what strategy would best manage it. This will enable them to engage in tension management fully and adequately. Our study results provide a good understanding of what learning activities or subprocesses ITO managers should pay more attention to, depending on whether the tensions are temporary (transient) or permanent (persistent).

Second, client and vendor managers should be aware of where and when their actions potentially contribute to the learning process for tension management in their ITO engagements. Our study highlights the role of individual partner resources and the learning paths for tension management. The study suggests that learning to manage transient tensions is informal, while learning to manage persistent tensions is formal, and that for either type of tension, the initial stages of learning are the responsibilities of both partners while enforcing the learning at the later stages is more the client's responsibilities. Managers should be aware that this learning is not a one-off process but can be recurrent depending on the type of tension experienced. This is because some tensions can be persistent; learning to manage tensions can lead to new tensions, and improperly interpreting or managing existing tensions might escalate the tensions. To better handle these, our study provides practitioners with insights into when existing tensions can escalate, when new tensions might arise, and where to lay emphasis, depending on the type of tension in question.

Study Limitations

Our study also has some limitations, which open up more avenues for future research. First, the study used an in-depth case study approach as a research strategy, which means its generalizability may be limited (Yin, 2018). Although the discussion unveils case-specific insights into the learning dynamics (such as different learning paths depending on whether tensions are transient or persistent), these insights provide a valuable foundation for future research to ensure the statistical generalizability of our findings. Future research could, for example, use this learning path dependency on the type of tension to develop and test hypotheses using survey data.

Second, while our study focuses on learning to manage tensions, it does not go deeper into emphasizing the interplay between tensions. Our research suggests that partners experienced multiple tensions simultaneously and that new tensions (e.g., learning from experience vs. learning from codified knowledge) can arise, indicating a cumulativeness of tensions (Ngah et al., 2023). These tensions, which could be knotted and interwoven, could present a different challenge than if they occurred independently. To this end, future research could, for example, build on the results of the current study to understand how firms learn to manage knotted and interwoven tensions in ITO engagements.

Third, our study results did not reveal any learning subprocess that aligns with knowledge institutionalization. This might have resulted from the nature of the case and the learning thereof, i.e., efficiency-driven learning. However, prior research has suggested that knowledge institutionalization is important for reasons such as efficient decision-making, reducing knowledge loss, and enhancing collaboration (Wiseman, 2010; Seo & Park, 2022). To explore our limitation, future research could, for example, use a multiple case study to investigate how contextual factors, such as the nature of ITO engagements, impact the institutionalization of knowledge on tension management and how to institutionalize this knowledge in such engagements.

Apart from these limitations, our case study has shown that an interplay (including feedback loops and governance) of activities pertaining to *experiencing tensions*, *sharing experiences*, *comprehending tensions*, *memorizing tensions*, and *addressing tensions* help ITO partner firms know how to deploy strategies to manage transient and persistent tensions in their engagements.

REFERENCES

Akkermans, H., Van Oppen, W., Wynstra, F., & Voss, C. (2019). Contracting outsourced services with collaborative key performance indicators. *Journal of Operations Management*, *65*(1), 22–47.

Ali, S., & Khan, S. U. (2016). Software outsourcing partnership model: An evaluation framework for vendor organizations. *Journal of Systems and Software*, *117*, 402–425.

Annosi, M. C., Martini, A., Brunetta, F., & Marchegiani, L. (2020). Learning in an agile setting: A multilevel research study on the evolution of organizational routines. *Journal of Business Research*, *110*, 554–566.

Argote, L., & Hwang, E. H. (2016). Organizational learning. In M. Augier & D. J. Teece (Eds.), *The Palgrave Encyclopedia of Strategic Management*. London: Palgrave Macmillan UK, 1–5.

Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science*, *22*(5), 1123–1137.

Argyris, Y. ‘Anna,’ & Ransbotham, S. (2016). Knowledge entrepreneurship: Institutionalising wiki-based knowledge-management processes in competitive and hierarchical organisations. *Journal of Information Technology*, *31*(2), 226–239.

Aubert, B. A., Kishore, R., & Iriyama, A. (2015). Exploring and managing the “innovation through outsourcing” paradox. *Journal of Strategic Information Systems*, *24*(4), 255–269.

Bahli, B., & Rivard, S. (2013). Cost escalation in information technology outsourcing: A moderated mediation study. *Decision Support Systems*, *56*, 37–47.

Bapna, R., Gupta, A., Ray, G., & Singh, S. (2016). IT outsourcing and the impact of advisors on clients and vendors. *Information Systems Research*, *27*(3), 636–647.

Beer, M., Voelpel, S. C., Leibold, M., & Tekie, E. B. (2005). Strategic management as organizational learning. *Long Range Planning*, *38*(5), 445–465.

Bell, J. E., Bradley, R. V., Fugate, B. S., & Hazen, B. T. (2014). Logistics information system evaluation: Assessing external technology integration and supporting organizational learning. *Journal of Business Logistics*, *35*(4), 338–358.

Bibbo, D., Michelich, J., Sprehe, E., & Lee, Y. E. (2012). Employing wiki for knowledge management as a collaborative information repository: An NBC universal case. *Journal of Information Technology Teaching Cases*, *2*(1), 17–28.

Brandl, K., Jensen, P. D. O., & Lind, M. J. (2018). Advanced service offshore outsourcing: Exploring the determinants of capability development in emerging market firms. *Global Strategy Journal*, *8*(2), 324–350.

Bui, Q. N., Leo, E., & Adelakun, O. (2019). Exploring complexity and contradiction in information technology outsourcing: A set-theoretical approach. *Journal of Strategic Information Systems*, *28*(3), 330–355.

Cao, L., Mohan, K., Ramesh, B., & Sarkar, S. (2013). Evolution of governance: Achieving ambidexterity in IT outsourcing. *Journal of Management Information Systems*, *30*(3), 115–140.

Carmeli, A., & Dothan, A. (2017). Generative work relationships as a source of direct and indirect learning from experiences of failure: Implications for innovation agility and product innovation. *Technological Forecasting and Social Change*, *119*, 27–38.

Chakrabarty, S., & Whitten, D. (2011). The sidelining of top IT executives in the governance of outsourcing: Antecedents, power struggles, and consequences. *IEEE Transactions on Engineering Management*, *58*(4), 799–814.

Chen, J., McQueen, R. J., & Sun, P. Y. T. (2013). Knowledge transfer and knowledge building at offshored technical support centers. *Journal of International Management*, *19*(4), 362–376.

Chua, A. L., & Pan, S. L. (2008). Knowledge transfer and organizational learning in IS offshore sourcing. *Omega-International Journal of Management Science*, *36*(2), 267–281.

Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. *The Academy of Management Review*, *24*(3), 522.

Dekker, H. C., & Van den Abbeele, A. (2010). Organizational learning and interfirm control: The effects of partner search and prior exchange experiences. *Organization Science*, *21*(6), 1233–1250.

Du, W. (Derek), & Pan, S. L. (2013). Boundary spanning by design: Toward aligning boundary-spanning capacity and strategy in IT outsourcing. *IEEE Transactions on Engineering Management*, *60*(1), 59–76.

Duhamel, F. B., Gutierrez-Martinez, I., Picazo-Vela, S., & Luna-Reyes, L. (2018). Determinants of collaborative interfaces in public-private IT outsourcing relationships. *Transforming Government- People Process and Policy*, *12*(1), 61–83.

Echajari, L., & Thomas, C. (2015). Learning from complex and heterogeneous experiences: The role of knowledge codification. *Journal of Knowledge Management*, *19*(5, SI), 968–986.

Edmondson, A. C. (2011). Strategies for learning from failure. *Harvard Business Review*, *89*(4), 48–57.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.

Ejodame, K., & Oshri, I. (2018). Understanding knowledge re-integration in backsourcing. *Journal of Information Technology*, *33*(2), 136–150.

Flores, L. G., Zheng, W., Rau, D., & Thomas, C. H. (2012). Organizational learning: Subprocess identification, construct validation, and an empirical test of cultural antecedents. *Journal of Management*, *38*(2), 640–667.

Geppert, M. (2005). Competence development and learning in British and German subsidiaries of MNCs: Why and how national institutions still matter. *Personnel Review*, *34*(2), 155–177.

Ghosh, B., & Scott, J. E. (2009). Relational alignment in offshore IS outsourcing. *MIS Quarterly Executive*, *8*(1), 19–29.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.

Gopal, A., & Gosain, S. (2010). The role of organizational controls and boundary spanning in software development outsourcing: Implications for project performance. *Information Systems Research*, *21*(4), 960–982.

Gozman, D., & Willcocks, L. (2019). The emerging Cloud Dilemma: Balancing innovation with cross-border privacy and outsourcing regulations. *Journal of Business Research*, *97*, 235–256.

Gupta, A. K., Smith, K. G., & Shalley, C. E. (2006). The interplay between exploration and exploitation. *Academy of Management Journal*, *49*(4), 693–706.

Handley, S., Skowronski, K., & Thakar, D. (2022). The single-sourcing versus multisourcing decision in information technology outsourcing. *Journal of Operations Management*, *68*(6–7), 702–727.

Heiskanen, A., Newman, M., & Eklin, M. (2008). Control, trust, power, and the dynamics of information system outsourcing relationships: A process study of contractual software development. *Journal of Strategic Information Systems*, *17*(4), 268–286.

Holmqvist, M. (2004). Experiential learning processes of exploitation and exploration within and between organizations: An empirical study of product development. *Organization Science*, *15*(1), 70–81.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science*, *2*(1), 88–115.

Jarzabkowski, P., Bednarek, R., Chalkias, K., & Cacciatori, E. (2022). Enabling rapid financial response to disasters: Knotting and reknotting multiple paradoxes in interorganizational systems. *Academy of Management Journal*, *65*(5), 1477–1506.

Jones, O., & Macpherson, A. (2006). Inter-organizational learning and strategic renewal in SMEs. *Long Range Planning*, *39*(2), 155–175.

Kale, P., & Singh, H. (2007). Building firm capabilities through learning: The role of the alliance learning process in alliance capability and firm-level alliance success. *Strategic Management Journal*, *28*(10), 981–1000.

Karimi-Alaghehband, F., & Rivard, S. (2020). IT outsourcing success: A dynamic capability-based model. *Journal of Strategic Information Systems*, *29*(1), 101599.

Kern, T., & Willcocks, L. (2002). Exploring relationships in information technology outsourcing: The interaction approach. *European Journal of Information Systems*, *11*(1), 3–19.

Kim, S., & Chung, Y. (2003). Critical success factors for IS outsourcing implementation from an interorganizational relationship perspective. *Journal of Computer Information Systems*, *43*(4), 81–90.

Kim, S. M., Anand, G., Larson, E. C., & Mahoney, J. (2019). Resource co-specialization in outsourcing of enterprise systems software: Impact on exchange success and firm growth. *Journal of Science and Technology Policy Management*, *10*(5), 1015–1046.

Koo, Y., Lee, J.-N., Heng, C. S., & Park, J. (2017). Effect of multi-vendor outsourcing on organizational learning: A social relation perspective. *Information & Management*, *54*(3), 396–413.

Kotlarsky, J., Oshri, I., Dibbern, J., & Mani, D. (2020). IS Sourcing. *MISQ Research Curation*, 18.

Kranz, J. (2021). Strategic innovation in IT outsourcing: Exploring the differential and interaction effects of contractual and relational governance mechanisms. *Journal of Strategic Information Systems*, *30*(1), 101656.

Kyndt, E., & Baert, H. (2013). Antecedents of employees' involvement in work-related learning: A systematic review. *Review of Educational Research*, *83*(2), 273–313.

Lacity, M., & Willcocks, L. (1998). An empirical investigation of information technology sourcing practices: Lessons from experience. *MIS Quarterly*, *22*(3), 363–408.

Lampel, J., Shamsie, J., & Shapira, Z. (2009). Experiencing the improbable: Rare events and organizational learning. *Organization Science*, *20*(5), 835–845.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710.

Leitch, R., & Day, C. (2000). Action research and reflective practice: Towards a holistic view. *Educational Action Research*, *8*(1), 179–193.

Lengnick-Hall, C. A., & Inocencio-Gray, J. L. (2013). Institutionalized organizational learning and strategic renewal: The benefits and liabilities of prevailing wisdom. *Journal of Leadership & Organizational Studies*, *20*(4), 420–435.

Levinthal, D. A., & March, J. G. (1993). The myopia of learning. *Strategic Management Journal*, *14*, 95–112.

Lioliou, E., Willcocks, L., & Liu, X. (2019). Researching IT multi-sourcing and opportunistic behavior in conditions of uncertainty: A case approach. *Journal of Business Research*, *103*, 387–396.

Lundvall, B.-Å. (2016). *The learning economy and the economics of hope*. Anthem Press.

Madsen, P. M., & Desai, V. (2010). Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Academy of Management Journal*, *53*(3), 451–476.

Mani, D., & Barua, A. (2015). The impact of firm learning on value creation in strategic outsourcing relationships. *Journal of Management Information Systems*, *32*(1), 9–38.

Markus, M. L., & Jacobson, D. D. (2015). The governance of business processes. In J. Vom Brocke & M. Rosemann (Eds.), *Handbook on Business Process Management 2*. Berlin, Heidelberg: Springer Berlin Heidelberg, 311–332.

Mihalache, M., & Mihalache, O. R. (2020). What is offshoring management capability and how do organizations develop it? A study of Dutch IT service providers. *Management International Review*, 37–67.

Miranda, S. M., & Kavan, C. B. (2005). Moments of governance in IS outsourcing: Conceptualizing effects of contracts on value capture and creation. *Journal of Information Technology*, *20*(3), 152–169.

Moe, N. B., Smite, D., Hanssen, G. K., & Barney, H. (2014). From offshore outsourcing to insourcing and partnerships: Four failed outsourcing attempts. *Empirical Software Engineering*, *19*(5), 1225–1258.

Ngah, E., Tjemkes, B., & Dekker, H. (2023). Relational dynamics in information technology outsourcing: An integrative review and future research directions. *International Journal of Management Reviews*, *00*, 1–28.

Nicholson, B., Babin, R., & Briggs, S. (2017). Exploring the effects of liminality on corporate social responsibility in inter-firm outsourcing relationships. *Journal of Information Technology*, *32*(1), 47–61.

Oshri, I., Kotlarsky, J., & Gerbasi, A. (2015). Strategic innovation through outsourcing: The role of relational and contractual governance. *Journal of Strategic Information Systems*, *24*(3), 203–216.

Park, J. Y., Im, K. S., & Kim, J. S. (2011). The role of IT human capability in the knowledge transfer process in IT outsourcing context. *Information & Management*, *48*(1), 53–61.

Pero, M., Rossi, M., & Terzi, S. (2018). Evolution of global product development networks: An exploratory study. *IEEE Transactions on Engineering Management*, *65*(1), 34–45.

Peukert, C. (2019). Determinants and heterogeneity of switching costs in IT outsourcing: Estimates from firm-level data. *European Journal of Information Systems*, *28*(3), 291–317.

Putnam, L. L., Fairhurst, G. T., & Banghart, S. (2016). Contradictions, dialectics, and paradoxes in organizations: A constitutive approach. *Academy of Management Annals*, *10*(1), 65–171.

Rodríguez, C., Langley, A., Béland, F., & Denis, J.-L. (2007). Governance, power, and mandated collaboration in an interorganizational network. *Administration & Society*, *39*(2), 150–193.

Seo, R., & Park, J.-H. (2022). When is interorganizational learning beneficial for inbound open innovation of ventures? A contingent role of entrepreneurial orientation. *Technovation*, *116*, 102514.

Sheep, M. L., Fairhurst, G. T., & Khazanchi, S. (2017). Knots in the discourse of innovation: Investigating multiple tensions in a reacquired spin-off. *Organization Studies*, *38*(3–4), 463–488.

Silva, M. M., Poleto, T., de Gusmao, A. P. H., & Costa, A. P. C. S. (2020). A strategic conflict analysis in IT outsourcing using the graph model for conflict resolution. *Journal of Enterprise Information Management*, *33*(6), 1581–1598.

Simons, R. (1994). How new top managers use control systems as levers of strategic renewal. *Strategic Management Journal*, *15*(3), 169–189.

Simons, R., & Dávila, A. (2021). How top managers use the entrepreneurial gap to drive strategic change. *European Accounting Review*, *30*(4), 583–609.

Smite, D., & Moe, N. B. (2020). Vendor switching: Factors that matter when engineers onboard their own replacement. *Journal of Systems and Software*, *169*.

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, *36*(2), 381–403.

Soderberg, A.-M., Krishna, S., & Bjorn, P. (2013). Global software development: commitment, trust and cultural sensitivity in strategic partnerships. *Journal of International Management*, *19*(4), 347–361.

Susarla, A. (2012). Contractual flexibility, rent seeking, and renegotiation design: An empirical analysis of information technology outsourcing contracts. *Management Science*, *58*(7), 1388–1407.

Teo, T. S. H., & Bhattacherjee, A. (2014). Knowledge transfer and utilization in IT outsourcing partnerships: A preliminary model of antecedents and outcomes. *Information & Management*, *51*(2), 177–186.

Tsang, E. (2002). Acquiring knowledge by foreign partners from international joint ventures in a transition economy: Learning-by-doing and learning myopia. *Strategic Management Journal*, *23*(9), 835–854.

Vicente, A. R. P., Ferasso, M., & May, M. R. (2018). Dynamic capabilities development and business model innovation: Evidences from IT industry in an emerging country. *International Journal of Business Innovation and Research*, *17*(2), 226–248.

Visnjic, I., Jovanovic, M., & Raisch, S. (2022). Managing the transition to a dual business model: Tradeoff, paradox, and routinized practices. *Organization Science*, *33*(5), 1964–1989.

Voss, C., Tsikriktsis, N., & Frohlich, M. (2002). Case research in operations management. *International Journal of Operations & Production Management*, *22*(2), 195–219.

Wang, M.-M., & Wang, J.-J. (2019). How vendor capabilities impact IT outsourcing performance: An investigation of moderated mediation model. *Journal of Enterprise Information Management*, *32*(2), 325–344.

Whitten, D., Chakrabarty, S., & Wakefield, R. (2010). The strategic choice to continue outsourcing, switch vendors, or backsource: Do switching costs matter? *Information & Management*, *47*(3), 167–175.

Wiener, M., Remus, U., Heumann, J., & Maehring, M. (2015). The effective promotion of informal control in information systems offshoring projects. *European Journal of Information Systems*, *24*(6), 569–587.

Wiseman, E. (2010). *The institutionalization of organizational knowledge: Learning to walk the talk*. Ottawa: Library and Archives Canada = Bibliothèque et Archives Canada.

Yakhlef, A. (2009). Outsourcing as a mode of organizational learning. *Strategic Outsourcing: An International Journal*, *2*(1), 37–53.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (*6th ed.*). SAGE Publications, Inc.

Zollo, M., & Reuer, J. J. (2010). Experience spillovers across corporate development activities. *Organization Science*, *21*(6), 1195–1212.

Zollo, M., & Singh, H. (2004). Deliberate learning in corporate acquisitions: Post-acquisition strategies and integration capability in U.S. bank mergers. *Strategic Management Journal*, *25*(13), 1233–1256.

Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science*, *13*(3), 339–351.

DECISION SCIENCES INSTITUTE

How Organizational Culture Affects Project Performance: US-Thailand Comparison

Qiannong Gu
Ball State University
Email: qgu@bsu.edu

Jie Yang
University of Houston - Victoria
Email: YangJ2@uhv.edu

Yixiu Yu
Ball State University
Email: yyu4@bsu.edu

Lu Wang
Ball State University
Email: wang@bsu.edu

ABSTRACT

With globalization advancing rapidly, encountering challenges related to national cultural differences and their impact on project management is unavoidable in multinational companies and partnerships involving diverse cultural backgrounds. This study aimed to survey project management professionals in both the United States and Thailand to investigate the influence of success factors in project management (such as tools, best practices, and managerial support) on project performance at individual, team, and project levels in these two countries. By comparing the results of confirmatory factor analysis on data collected from the two countries, we discovered that while all success factors improved performance across all levels in both countries, the specific combinations of success factors and organizational culture components had different effects on project performance between the two countries. Our findings provide actionable insights for project managers working with culturally diverse teams, empowering them to implement suitable success factors and enhance targeted performance. Additionally, the research methodology and approach proposed in this study can be applied to explore project management practices in different countries and settings, further advancing our understanding of the role of culture in project performance.

KEYWORDS: Project management, National culture, Organizational culture, Success factors, Confirmatory factor analysis

INTRODUCTION

Given the rapid changes in a globalized business environment, the collaboration between project team members across countries and cultures are inevitable and becomes more and more popular and challenging. Understanding the culture difference and its impact on project performance is crucial for the success of a project which involves different cultural elements. Most existing research assume the homogeneousness of project team members. The conclusions and results are most likely valid only for a certain culture which is investigated. However, this assumption may put some constraints on the project management knowledge from the literature while project managers implement them in practice. A successful approach in one culture may not succeed in another. For example, a common project management practice in a well-developed western country, such as the United States, may not be applicable in an eastern developing country, such as Thailand, vice versa.

This study surveyed project professionals in the United States and Thailand and investigated how eastern and western cultural backgrounds will influence the ways that project performances are evaluated. For example, Thailand is the only country in the southeast Asia that has escaped colonial rule. Buddhist religion, the monarchy, and the military have helped to shape its society and politics. Because of its geographical location as Golden Peninsular separating the Pacific Ocean and Indian Ocean, Thailand has become an important economic powerhouse bridging between Eastern and Western nations. Unlike some other Asian nations whose strong traditions and cultures are transfixed with the society, Thailand and Thai are very flexible and open to adopt new traditions and cultures.

For business ventures, Thailand has adopted many western business paradigms from the United States. Various modern business concepts and methodologies are taught in business schools in Thailand. Project management courses are offered commonly at the undergraduate and graduate levels. Thai business schools have produced many knowledgeable project professionals and managers. An active Thai's Project Management Professional (PMP) Association is a strong indication that Thai business managers recognized that project management is crucial for the success of project implementation. Although Thailand has adopted western project management methodologies and business paradigms, many concepts and techniques are not well regarded and the success rate for the implementation is still questionable. The Thailand is still steeped in old traditions. Thai are more attuned to their cultural Buddhism traditions. They respect their peers and value seniority beyond business interactions and value and prefer indirect and soft spoken. Therefore, many practices in project management that require direct communication are not well received in Thai society. Because of the cultural differences between the United States and Thailand, it would be interesting to study, 1) how national culture background affects the project management best practices, tools, and management support, and 2) how the national cultural difference affects project management performances of individuals, team, and whole project.

LITERATURE REVIEW

The extant literature reveals numerous factors that influence the successful implementation of projects in organizations. Besteiro *et al.* (2015) described *scope, time, and cost* as the iron triangle (also called triple constraints) for a traditional view on project success. Fortune *et al.* (2013) stated that the project success factors should be part of the organizations' strategic perspective and several influences are derived from the expectations of the stakeholders. Other studies agreed that all projects must have contingency plans set in place (Lundin and Soderholm, 1995; Soderlund, 2002; Carvalho and Rabechini, 2010). Fortune and White (2006) analyzed 63 publications and identified that the most important critical success factors are clear goals, senior executive support, and adequate resources. Hyvrari (2006) examined the critical factors and the failures in project management. The success factors include clear objectives, commitment to the end user, adequate resources, ability to coordinate, effective leadership, commitment and flexibility with resources, support from the upper management, clear job description, structuring by project, and technological and economic environment. The same study also showed that project communication, consulting the client, acceptance from the client, support from the upper management, project schedule, mission of the project, project execution, troubleshooting, staff management, and monitoring and control are listed as critical success factors. However, communication is the most important critical success factor in large companies. Besteirio, de Souza Pinto, and Novaski (2015) asserted that knowledge management, leadership, and experience support the project to be successful. Anderson *et al.* (2006) studied the relationship between critical factors and project success and proposed the success factors with three success scales: managerial delivery ability, impacts of the project, and experiences captures. They proposed nine critical success factors including communication, planning approval by stakeholders, formal and well-structured approach, commitment to the project, influence of stakeholders, understanding and accepting the proposal, restrictions, flexibility in the execution, and influence over the processes of the project. The results showed that communication is a significant contribution in order to establish a trusting relationship between the project team members. Besteiro *et al.* (2015) conducted an exploratory empirical research on project managers from 28 companies in order to classify critical success factors into 4 driver groups. They proposed 57 variables altogether with 18 for managerial liabilities group, 19 for critical success factors group, 13 for monitoring and control group and 7 for lessons learned.

The project management success factors are also reported in studies using students as a test subject. Using evaluative and developmental measures at the individual and team levels, Kemery and Stickney (2014) assessed a multifaceted, multilevel approach for acquiring and assessing teamwork knowledge, skills, and abilities in an undergraduate business course. Teamwork knowledge, individual teamwork behavior, and collaborative peer rating were the three assessments administered. Larson & Drexler (2010) implemented service-learning projects in a project management course to introduce project management concepts and to determine success factors. Using fund-raising projects to assess the hard and soft skill aspects, learning the project management framework, the work breakdown structure, planning, scheduling, and controlling projects are described as key essential ingredients for project success. Since students enrolled in the project management course were expected to master the principles and application of the project management framework, emphasis was placed on both the theory and the practice of project management. A key success factor was social capital, the ability to collaborate as a unified team and to use social networks to help facilitate successful project outcomes. Student teams learned how to merge their interests with those others and cultivate friendships. The importance of teamwork dominated the lessons-learned documentation.

In summary, a great deal of literature has addressed various project management success factors in organizational context. Many studies focused on creating the set of success factors (Lundin and Soderholm, 1995; Soderlund, 2002; Fortune and White, 2006; Hyvrari, 2006; Carvalho, 2010; Fortune *et al.*, 2013; and Besteiro *et al.*, 2015) and others focused on establishing a relationship between the use of the success factors and project performance (Zmud, 1980; McFarlan, 1981; Henderson and Lee, 1992; Settle-Murphy and Thornton, 1999; Jiang et al., 2000; Yetton et al., 2000; Leung, 2001; Anderson et al, 2006; Martin et al., 2007). However, there is few research in reference to project management and cultural differences between nations. Since the project management discipline contains a social science component in which the complexity and difficulty of a project are introduced by human behaviors, it creates a research gap to explore how the differences in national culture affect the ways project managers and team members choose different best practices, tools, and management support to achieve different levels of project performance. National culture background is treated as a moderating variable. Some pioneers paved the way for further investigation and exploration. Gu *et al*. (2014) studied the moderation effects of organizational culture and environmental pressures on IT project performance. They collected data from United States and China. Keil *et al*. (2000) conducted experiments in three national cultures (Finland, The Netherland, and Singapore) and revealed that there is no difference for decision makers to decide the continuity of a software project across cultures.

The purpose of this study is threefold. First, there are various factors hindering the progress of projects. Some are procedural in nature; however, many are human related. It would be beneficial to identify and catalog the hindering factors and establish appropriate ways to alleviate them accordingly. This study attempts to evaluate project management success factors designed to guide project managers and team members while working on a project. Throughout this study, the term of success factors is used to refer to project management best practices. Second, once the success factors have been analyzed, the comparison study will commence. The relationship between success factors and project performances at individual, team, and project levels is examined for both the United States and Thailand professionals. The results are then compared. Third, the “organizational cultural differences” variable is established using three main sub-constructs (Organizational Behaviors, Organizational Relationships, and Organizational Resistance). The organizational cultural differences variable is then used as a moderating variable for the relationships between the success factors and project performances. Furthermore, it would be interesting to see if there is any difference in terms of the relationships between success factors and project performance under different national cultures, i.e., the U.S. and Thailand in this study. The organizational culture is assumed to be justified or influenced by national culture.

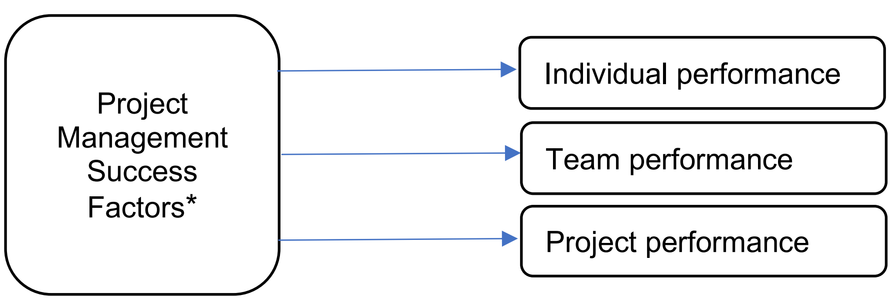
HYPOTHESES/MODEL

Theoretical Framework and Rationale

Initial preliminary interview phase results provide adequate rationale for the survey phase of the research study. The goal of this survey research is to explore specific project management issues across a wide variety of projects and project management courses to formulate specific project management success factors that contribute to project success. The identified project management success factors might provide guiding principles for both project managers and team members as they venture through phases of project implementation. Using the project management success factors, team members can learn to follow best practices and managers can successfully exhort project management knowledge.

Based on the theories of project management success factors presented in project management literature, three major hypotheses accompanied by supporting hypotheses have been developed. The three independent variables which might have an influence on project performance are project management (1) tools, (2) practices, and (3) support. The dependent variables are three project performance criteria (1) individual, (2) team, and (3) project. These are used for evaluating the success of project performance in this study. Project performance refers to a single project outcome. Team performance refers to a long-term performance of a team consistent of relatively stable team members. It is normally based on the outcomes of multiple projects. Figure 1 shows a general project management framework for success factors and performance. The next section will discuss individual hypothesis development in detail which is accompanied by supporting hypothesized framework.

Figure 1: A Project Management Framework for Success Factors and Performance



*Project Management Success Factor = {tools; practice; support}

Organizational Cultural Differences

Leadership, internal culture, and resistance are used to assess the organizational cultural differences variable in this study. In the western culture, common organizational practices focus on individualism, not collectivism (Napier *et al.*, 2007; Sauer and Horner-Reich, 2009). Leadership is a skill acquired after several years of team-based or workplace experiences; however, exposure to leadership roles is deficient for some project managers. They are unable to communicate effectively, delegate tasks, discipline team members, or demand respect. Alternatively, peer acceptance is viewed as an important trait among team members; therefore, it is critical to assert authority among peers for the success of a project. The lack of motivation and coordination typically erode the benefits of collaboration or leadership development.

Project success is often determined by its weakest part. One incompetent team member could disrupt the entire team. Differing viewpoints lead to team conflict, unproductivity, and mistrust. Many members have not yet garnered an appreciation of the power of a support system within a team. Establishing teams as a system with interacting components that aid each other throughout the project life cycle is paramount in achieving peak team performance, but frequently not as important for members who earn an individual performance. Promoting the importance of a supportive team environment is a challenge for organizations.

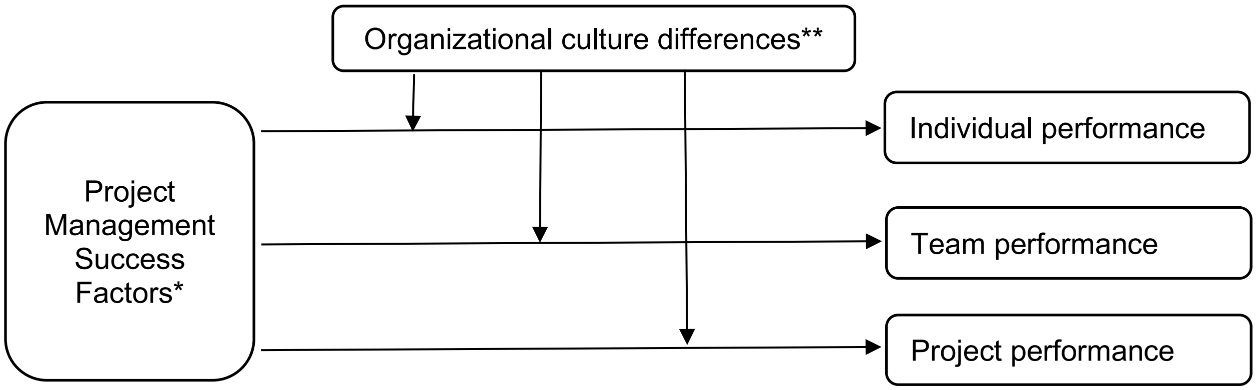
According to the *PMBOK*® Guide, project management knowledge areas, processes, tools, and techniques demand much more planning than general management (PMI, 2021). Basic project management tools are tangible items used to perform project activities. Physical documentation such as, the project charter, project scope statement, communications management plan, work breakdown structure (WBS), earned value management (EVM), critical path analysis, decisions tree, and risk assessments are valuable tools that contribute to a project's outcome. Larson and Drexler (2010) found that after a 10-week project management course that required the documentation of a service-learning project, students appeared to have a better grasp of how to use project management tools necessary for the satisfactory completion of a project. Inimitable proprietary tangible tools were found to contribute to both project-level and firm-level performance, and inimitable embedded intangible tools were also found to contribute to firm-level performance (Mathur *et al.*, 2014). Based on these considerations, the following set of hypotheses is proposed:

H1a: Organizational cultural differences mitigate the relationship between project management tools success factor variable and individual performance.

H1b: Organizational cultural differences mitigate the relationship between project management tools success factor variable and team performance.

H1c: Organizational cultural differences mitigate the relationship between project management tools success factor variable and project performance.

Figure 2: Project Management Tools/Practices/Support Success Factor and Performance



*Project Management Success Factor = {tools; practice; support}
**Organizational culture difference = {leadership; internal culture; resistance}

Best Practices

Project management practices are a combination of establishing a detailed work plan that manages scope, time and cost, identifies deliverables, and assesses risks. Clarifying scope by establishing transparent goals and objectives and controlling for scope creep are paramount to project success. Equally important is the implementation of effective communication channels for all vested stakeholders. Ofori (2013) found that the critical factors that contribute to the success of a project include top management support, effective communication, clarity of project purpose and goals, and stakeholder involvement. This finding was supported by Davidson (2002) and Meredith *et al.* (2015) which emphasized the employment of best practices to ensure project outcomes. Based upon these considerations, the following set of hypotheses is proposed:

H2a: Organizational cultural differences mitigate the relationship between project management practices success factor variable and individual performance.

H2b: Organizational cultural differences mitigate the relationship between project management practices success factor variable and team performance.

H2c: Organizational cultural differences mitigate the relationship between project management practices success factor variable and project performance.

Organizational support refers to employees' perception concerning the extent to which the organization values their contribution and cares about their well-being. Organizational support has been found to have important consequences employee performance (Eisenberger, 1986). Organizational support from upper management, business partners, staff personnel, and team members is a critical success project management factor (Verburg *et al.*, 2013; Yazici, 2009). Organizational support plays a key role in project team success by providing important administrative assistance (Thamhain, 1990) and by aligning human resource management systems with the project needs (Pinto and Slevin, 1987a, b; Thamhain and Wilemon, 1987; Brown and Eisenhardt, 1995). Project outcomes and organizational support are closely aligned and play a major role in project success (Gray, 2001). Organizational culture (values, beliefs, and behavioral norms) positively influences project success. However, drawing on extensive field research involving project management professionals in major British organizations, Gray found project success is shown to decline as the level of personal and environmental threat perceived by project staff increases. In addition, organizational change and conflict are negatively associated with project success. Scott-Young and Samson (2007) examined project cost, schedule, and operability and suggested micro-management or interference by strong senior management may negatively impact project schedules. As a result of empirical evidence that purports organizational support relates to individual, team, and project performance, we propose the third set of hypotheses:

H3a: Organizational cultural differences mitigate the relationship between project management support success factor variable and individual performance.

H3b: Organizational cultural differences mitigate the relationship between project management support success factor variable and team performance.

H3c: Organizational cultural differences mitigate the relationship between project management support success factor variable and project performance.

METHODS

Instrument development for the success factors and project management performance was carried out in three phases: (1) item generation, (2) online data collection study, and (3) data analysis. An extensive literature review was conducted to identify the domain of constructs and generate the initial measurement items. Next, a questionnaire was developed and delivered to students through online survey online platform called Qualtrics. Third, a data set was collected and analyzed to validate the instrument and test the hypothesized model.

Item Generation

Items for each construct were developed based on extensive literature reviews and brain storming with multiple researchers in the project management field. All items are measured on a six-point Likert type scales where 1 = “Strongly Disagree”, 2 = “Disagree”, 3 = “Somewhat Disagree”, 4 = “Somewhat Agree”, 5 = “Agree”, and 6 = “Strongly Agree”. After the rigorous sorting procedures were done, three sub-constructs for the success factors and three sub-constructs for project management performance were emerged. Table 1 lists all items in the initial measurement models for the success factors construct. There are 7 items for Project Management Tools construct, 6 items for Project Management Practices construct, and 6 items for Project Management Support construct.

Table 1: Items for Project Management Success Factors Construct

PROJECT MANAGEMENT SUCCESS FACTORS
The following situations describe the extent to which the **project management team is given sufficient project management tools, project management practices, and project management support to carry out a successful project implementation**. Please circle the appropriate number to indicate the extent to which you agree or disagree with each statement as applicable to your project.
PROJECT MANAGEMENT TOOLS - we use the project management tools to achieve the following project management activities.
Communicating among team members (Performance reports)
Sharing information
Making decisions (Decision tree)
Evaluating project progress (Critical path)
Balancing budget (Earned Value Management-EVM)
Informing the evolution of the project (Gantt chart)
Clear detailed task list (Work Break Structure-WBS)
PROJECT MANAGEMENT PRACTICES - we identify the following activities as key factors for project success.
Clear scope definition of project
Establishing goals and deliverables
Commitment of stakeholders, decision-makers, project managers and team members
Clear time requirement
Clear resource and budget requirements
Clear contingency plan
PROJECT MANAGEMENT SUPPORT - we receive help from the following supporters.
Upper management
Business partners
IT personnel
Project managers
Team members
Managerial support for project management processes

Table 2 lists all items in the initial measurement models for project management performance construct. There are 5 items for Individual Performance construct, 6 items for Team Performance construct, and 5 items for Project Performance construct.

Table 2: Items for Project Management Performance Construct

PROJECT MANAGEMENT PERFORMANCE
The following statements describe typical **project management performance objectives**. Please circle the appropriate number to indicate the extent to which you agree or disagree with each statement as applicable to your project.
INDIVIDUAL PERFORMANCE - we use the following criteria to measure team members' performance.
Achieve the scope objectives
Meet time requirements
Manage cost effectively
Sense of accomplishment
Quality of work
TEAM PERFORMANCE - we set the following criteria to measure team performance.
Achieve the scope objectives
Meet time requirements
Manage cost effectively
Team dynamics
Alignment of project activities (for example, the outcomes from one activity should be tightly matched with the project's objectives)
Balance of obligation
PROJECT PERFORMANCE - we use the following criteria to measure project performance.
Achieve the scope objectives
Meet time requirements
Manage cost effectively
Quality of deliverables
Customer satisfaction

Table 3 lists all items in the initial measurement models for Cultural Differences construct. There are 5 items for Leadership construct, 5 items for Internal Culture construct, and 5 items for Organizational Resistance construct.

Table 3: Items for Cultural Differences Construct

CULTURAL DIFFERENCES
The following statements describe typical relationships among project management team members. Please circle the appropriate number that best indicates the level of your stakeholder relationships, internal culture, and organizational resistance.
LEADERSHIP **In our organization, we set the following criteria to measure leadership.**
Create a clear vision
Initiate an atmosphere of mutual trust
Take responsibility for decisions without shifting blame
Embrace authentic diversity
Celebrate organizational and team success
INTERNAL CULTURE **In our organization, we set the following criteria to measure internal culture.**
Awareness of project management value among all staff
Reward collaboration among team members
Create an open environment
Establish specific expectations
Adopt continuous improvement
ORGANIZATIONAL RESISTANCE **In our organization, we set the following criteria to measure business resistance**
Lack of trust due to internal politics
Lack of trust due to external politics
Lack of acceptance due to cultural differences
Lack of acceptance due to opposition of changes
Lack of understanding of the usefulness of formal project management methodologies

Survey Methods, Data Collection, and Sample Characteristics

A cross-sectional self-administered online survey was conducted. The sample frame was obtained from 121 college students enrolled in project management classes during the Spring semester. Out 121 students who received the invitation to participate in the survey, 94 completed the survey (77.69 % respond rate). All three sections were taught by two professors and the courses were mixed between the online (10%) and on-campus (90%) students. One of the key requirements is for students to complete the final team project. A team of 5-6 students worked on a creative project with the outside community from the initiation phase until reaching the project close-out phase. The examples of project range from creating a fund-raising event for a close-by nursing home to running a competitive baseball game for donations. Table 4 shows sample demographic information.

Table 4: Sample characteristics

1	Role in the project team		5	Number of members in the project team	
	Project manager	13 %		1	1%
	Member	48 %		2-4	57%
	Team leader	22 %		5-10	41%
	Organizer	17 %		Over 10	0%
2	Highest level of education				
	High school	56 %	6	Mode of content delivery	
	Two-year college	22 %		On campus	87%
	Bachelor's degree	28 %		Online	13%
	Master's degree	1 %			
	Doctor's degree	0 %	7	Nature of project collaboration	
	Other	2 %		On campus	90%
3	Important skills for the project success			Online	10%
	Leadership	68 %			
	Responsibility	67 %	8	Primary use of technology applications	
	Teamwork	84 %		E-mail	88 %
	Resources	27 %		Word processing	67 %
	Clear goals	66 %		Spread sheet	80 %
	Communication	81 %		Database	30 %
	Reward	15 %		Programming tools	17 %
	Punishment	7 %		Project management software	44 %
4	Length of project in days			Social media	22%
	1-5	3 %		Cloud space	37 %
	5-10	11 %		Others	18 %
	20-30	13 %			
	30-50	21 %			
	Over 50	52 %			

To establish content validity, items for the various constructs were reviewed by three academicians. Based on the feedback from the academicians, ambiguous items were either modified or eliminated. New items were added when necessary. A confirmatory factor analysis (CFA) using structural equation modeling method in Analysis of Moment Structures (AMOS) was conducted to refine the measurement models (Hair et al., 2006).

For the instrument refinement process, the data collected in the US are used to modify th instrument, when then be used to collect data from Thailand. The final 94 US responses were used for further data analysis. As per the guidelines of Bagozzi (1980), the important properties for a measure to be reliable and valid include content validity, internal consistency of operationalization (unidimensionality and reliability), and construct validity (discriminant and convergent). Content validity was determined through a comprehensive review of the literature, pilot tests, assessment by a panel of practitioners, and academics to ensure that measurement items covered the domain of the variable (Nunnally, 1978; Churchill, 1979). Convergent validity criteria requires that there be one single latent variable underlying a set of measurement items (Anderson and Gerbing, 1988). The degree of convergent validity is tested through CFA (using SEM) to assess the measurement model properties indicated by the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI). The GFI indicates the relative amount of variance and covariance jointly explained by the model. The AGFI differs from the GFI in adjusting for the number of degrees of freedom (Byrne, 1989). Both range from 0 to 1. Values of 0.90 or more are considered a good fit (Hair et al., 1998). The normed fit index (NFI) and comparative fit index (CFI) indicate a relative comparison of the proposed model to the null model. NFI and CFI's values above 0.90 are acceptable (Hu and Bentler, 1999). The next set of fit statistics focus on the root mean square error of approximation (RMSEA). The RMSEA considers the error of approximation and is expressed per degree of freedom, thus making the index sensitive to the number of estimated parameters in the model; values less than 0.05 indicate good fit, values up to 0.08 represent reasonable errors of approximation (Browne and Cudeck, 1993), values up to 0.10 indicate mediocre fit, and those greater than 0.10 indicate poor fit (MacCallum *et al.*, 1996). The original measurement items were tested according to these indices and items contributing to an unacceptable model fit were deleted. Items in *italic* were deleted due to low factor loadings or error term correlations. Cronbach's alpha was also used to assess the reliability of each construct. Values of 0.70 and higher are considered acceptable (Hair *et al.*, 2006). Tables 5, 6, and 7 list the items shown in the final measurement models' result. The results show that all constructs are valid and reliable according to the minimum requirements. A cross-sectional self-administered online survey was conducted through the networking/connection of an administrator at the Thai PMP Association. Out of 32 responses, 27 surveys are completed and usable (84.38% respond rate).

Table 5: Model Fit Indices and Reliabilities for Project Management Success Factors

PROJECT MANAGEMENT TOOLS	Final Model Fit
Communicating among team members (Performance reports)	GFI = 0.98 AGFI = 0.92 CFI =0.99 NFI = 0.98 RMSEA = 0.06 α = 0.82
Sharing information	
Making decisions (Decision tree)	
Evaluating project progress (Critical path)	
Balancing budget (Earned Value Management-EVM)	
Informing the evolution of the project (Gantt chart)	
Clear detailed task list (Work Break Structure-WBS)	
PROJECT MANAGEMENT PRACTICES	Final Model Fit
Clear scope definition of project	GFI = 0.99 AGFI =0 .93 CFI = 1.00 NFI = 0.99 RMSEA =0.05 α = 0.87
Establishing goals and deliverables	
Commitment of stakeholders, decision-makers, project managers and team members	
Clear time requirement	
Clear resource and budget requirements	
Clear contingency plan	
PROJECT MANAGEMENT SUPPORT	Final Model Fit
Upper management	GFI = 1.00 AGFI =0.96 CFI = 1.00 NFI =0.99 RMSEA =0.00 α = 0.79
Business partners	
IT personnel	
Project managers	
Team members	
Managerial support for project management processes	

Table 6: Model Fit Indices and Reliabilities for Project Management Performance

INDIVIDUAL PERFORMANCE	Final Model Fit
Achieve the scope objectives	GFI = 1.00; AGFI = 1.00 CFI = 1.00 NFI = 1.00 RMSEA = .00 α = 0.85
Meet time requirements	
Manage cost effectively	
Sense of accomplishment	
Quality of work	
TEAM PERFORMANCE	Final Model Fit
Achieve the scope objectives	GFI = 0.98; AGFI = 0.93 CFI = 1.00 NFI = 0.97 RMSEA = 0.03 α = 0.82
Meet time requirements	
Manage cost effectively	
Team dynamics	
Alignment of project activities (for example, the outcomes from one activity should be tightly matched with the project's objectives)	
Balance of obligation.	
PROJECT PERFORMANCE	Final Model Fit
Achieve the scope objectives	GFI = 0.98; AGFI = 0.89 CFI = 0.98; NFI = 0.97 RMSEA = 0.11 α = 0.79
Meet time requirements	
Manage cost effectively	
Quality of deliverables	
Customer satisfaction.	

Table 7: Model Fit Indices and Reliabilities for Cultural Differences

LEADERSHIP	Final Model Fit
Create a clear vision	GFI = .94 AGFI = .78 CFI = 1.00 NFI = .97 RMSEA = .08 α = .96
Initiate an atmosphere of mutual trust	
Take responsibility for decisions without shifting blame	
Embrace authentic diversity	
Celebrate organizational and team success	
INTERNAL CULTURE	Final Model Fit
Awareness of project management value among all staff	GFI = .96 AGFI = .84 CFI = 1.00 NFI = .98 RMSEA = .00 α = .96
Reward collaboration among team members	
Create an open environment	
Establish specific expectations	
Adopt continuous improvement	
Awareness of project management value among all staff	
ORGANIZATIONAL RESISTANCE	Final Model Fit
Lack of trust due to internal politics	GFI = .94 AGFI = .78 CFI = 1.00 NFI = .97 RMSEA = .04 α = .92
Lack of trust due to external politics	
Lack of acceptance due to cultural differences	
Lack of acceptance due to opposition of changes	
Lack of understanding of the usefulness of formal project management methodologies	

Results and Discussion

The results from the assessment of the structural model appear in Tables 8 and 9. All hypothesized relationships without moderation effects are statistically significant at $p<0.001$ level or better. Hypothesis 1a is supported ($\beta=0.62$, $t=4.12$), which suggests that Project Management Tools have a direct impact on Individual Performance. Hypothesis 1b is supported ($\beta=0.56$, $t=3.60$), which lends support that Project Management Tools have a direct impact on Team Performance. Hypothesis 1c is supported ($\beta=0.58$, $t=3.96$), which indicates that a high level of Project Management Tools used leads to a high level of Project Performance. Hypothesis 2a is supported ($\beta=0.77$, $t=6.41$) indicating high level of best Project Management Practices implemented to a high level of Individual Performance. Hypothesis 2b is supported ($\beta=0.66$, $t=5.06$). This suggests that a high level of best Project Management Practices adopted leads to a high level of increased Team Performance. Hypothesis 2c is supported ($\beta=0.56$, $t=4.93$), which lends support to the claim that Project Management Practices is positively related to Project Performance. Hypothesis 3a is supported ($\beta=0.92$, $t=3.50$), which implies that organizational support for project management may have a direct effect on individuals performing project management tasks. Hypothesis 3b is supported ($\beta=0.87$, $t=3.38$), which lends support to the claim that organizational support for project management is positively related to Team Performance. Finally, Hypothesis 3c is supported ($\beta=0.64$, $t=3.72$), which implies that Project Management Support may have a direct effect on Project Performance.

We noticed that Thailand and US sample sizes are unequal (i.e. 27 and 92), so we performed random sampling to see whether the uneven sample sizes of the two groups would be a concern. We randomly selected 27 cases from the total 92 US sample. We then repeated the above analysis based on the equal sample sizes of Thailand and US groups (i.e., 27 and 27). We repeated random sampling procedure for multiple times, and we found that the results are consistent with our previous findings That is, the imbalanced sample sizes from two countries was not a concern in our study. The means of the coefficients from the random US samples can be found in the Table below.

Table 10: US random sampling results

	Mean of βs (US random samples)
Tools → Individual Performance	0.60
Tools → Team Performance	0.60
Tools → Project Performance	0.56
Practice → Individual Performance	0.64
Practice → Team Performance	0.46
Practice → Project Performance	0.38
Support → Individual Performance	0.52
Support → Team Performance	0.55
Support → Project Performance	0.33

DISCUSSION AND CONCLUSIONS

The main purpose of this study is to examine the effects of national culture on the causality of project management tools to project performances. We compare the results without moderation effects between two countries in Table 11. We find that there is no difference in terms of team performance between two countries. We observe the same results even with the consideration of moderation effects. Team performance is based on a long-term observation and evaluation.

Table 8: Hypothesis Testing Results for *US* Respondents

***US* Respondents**	***No Moderator***		***Moderator***					
			Leadership		***Internal Culture***		***Org. Resistance***	
Hypothesis	***β***	***t***	***β***	***t***	***β***	***t***	***β***	***t (p)***
H1a: PM Tools → Individual Performance	0.57	4.40***	0.63	6.55***	0.58	5.70***	0.13	1.11
H1b: PM Tools → Team Performance	0.61	4.34***	0.68	5.69***	0.63	5.24***	0.18	1.61
H1c: PM Tools → Project Performance	0.54	4.38***	0.58	5.79***	0.54	5.24***	0.22	2.00*
H2a: PM Practices → Individual Performance	0.65	6.21***	0.73	7.69***	0.73	7.39***	0.18	1.62
H2b: PM Practices → Team Performance	0.52	4.52***	0.71	6.08***	0.71	5.97***	0.21	1.85
H2c: PM Practices → Project Performance	0.42	4.41***	0.58	5.90***	0.56	5.62***	0.19	1.76
H3a: PM Support → Individual Performance	0.46	4.52***	0.67	7.09***	0.60	5.96***	0.14	1.23
H3b: PM Support → Team Performance	0.50	4.26***	0.71	6.08***	0.65	5.36***	0.20	1.76
H3c: PM Support → Project Performance	0.35	3.25***	0.54	5.44***	0.46	4.40***	0.14	1.27

***$p < .001$; **$p < 0.01$; *$p < 0.1$

Table 9: Hypothesis Testing Results for *Thai* Respondents

***Thai* Respondents**	***No Moderator***		***Moderator***					
			Leadership		***Internal Culture***		***Org. Resistance***	
Hypothesis	***β***	***t***	***β***	***t***	***β***	***t***	***β***	***t***
H1a: PM Tools → Individual Performance	0.70	3.32**	0.72	4.74***	0.75	5.10***	0.47	2.62**
H1b: PM Tools → Team Performance	0.59	2.57**	0.89	4.54***	0.89	4.50***	0.37	1.83
H1c: PM Tools → Project Performance	0.77	3.26***	0.88	4.53***	0.95	4.64***	0.43	2.10*
H2a: PM Practices → Individual Performance	0.60	3.61***	0.69	4.46***	0.74	5.02***	0.42	2.28*
H2b: PM Practices → Team Performance	0.64	3.35***	0.89	4.52***	0.93	4.87***	0.40	1.99*
H2c: PM Practices → Project Performance	0.61	3.26***	0.79	4.23***	0.87	4.74***	0.34	1.64
H3a: PM Support → Individual Performance	0.48	2.65**	0.66	4.18***	0.47	2.61**	0.37	1.98*
H3b: PM Support → Team Performance	0.59	2.86**	0.84	3.89***	0.37	1.83	0.38	1.85
H3c: PM Support → Project Performance	0.65	3.45***	0.74	3.82***	0.43	2.10*	0.36	1.75

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.1$

Therefore, it proves that the nationality difference will not affect the project management practice in a long run. However, we observed the significant difference between two countries on the effect of tools on individual performance and the effect of support on project performance. Those results reveal that nationality difference will impact the performance of a project in a short-run at the operational (individual and project) levels. Standardization of project management knowledge, education, and practice mitigates the cultural difference in a long run. Globalization of the business and supply chain management also help to fill the cultural gap.

Our research findings hold significant value for project managers who are responsible for leading teams comprising individuals with diverse national cultural backgrounds. By understanding the insights we have gained, these project managers can make informed decisions and implement appropriate success factors to enhance performance in their respective projects.

The first key insight derived from our study is the recognition of the critical role played by organizational culture in project performance. We have observed that the specific combination of success factors and components of organizational culture has a significant impact on project outcomes. By acknowledging and understanding these dynamics, project managers can proactively shape and align the organizational culture to support project success. This may involve fostering an environment of open communication, promoting cross-cultural understanding and collaboration, and adapting management practices to accommodate diverse cultural perspectives. Through such efforts, project managers can create a conducive atmosphere that maximizes the potential of diverse cultural backgrounds within their teams.

Furthermore, our research emphasizes the importance of tailoring success factors to suit the targeted performance goals. Project managers can leverage the knowledge gained from our study to identify the success factors that are most influential in their specific context. This understanding allows them to allocate resources, implement appropriate tools and practices, and provide the necessary managerial support to optimize project outcomes. By customizing their approach based on cultural nuances and specific project requirements, project managers can enhance team performance and achieve better project results.

Moreover, our research methodology and analytical approach can be extended to investigate project management practices and their effects on project performance in other countries and contexts. By applying similar research frameworks, project managers and researchers can gain deeper insights into the interplay between cultural factors and project outcomes across diverse settings. This expansion of research will contribute to a broader understanding of the influence of national culture on project management and enable the development of more robust strategies and practices for achieving project success globally.

In conclusion, our findings offer valuable insights and practical implications for project managers leading teams with diverse national cultural backgrounds. By leveraging these insights and implementing appropriate success factors, project managers can enhance targeted performance and overcome the challenges posed by cultural differences. Furthermore, our research opens avenues for further exploration, enabling project management professionals to investigate the effects of organizational culture and success factors in various countries and contexts, ultimately contributing to improved project outcomes on a global scale.

	US		Thailand		Difference
No Moderator	***β***	***t p***	***β***	***t p***	
Tools → Individual Performance	0.57	4.40 ***	0.70	3.32***	χ2 = 4.70*, p =0.03
Tools → Team Performance	0.61	4.34 ***	0.59	2.57**	χ2 = 1.67, p = 0.20
Tools → Project Performance	0.54	4.38 ***	0.77	3.26***	χ2 = 2.40, p = 0.12
Practice → Individual Performance	0.65	6.21 ***	0.60	3.65***	χ2 = 1.02, p = 0.31
Practice → Team Performance	0.52	4.52 ***	0.64	3.35 ***	χ2 = 2.28, p = 0.13
Practice → Project Performance	0.42	4.41***	0.61	3.26 ***	χ2 = .49, p = 0.48
Support → Individual Performance	0.46	4.52 ***	0.48	2.65 **	χ2 = 1.16, p = 0.28
Support → Team Performance	0.50	4.26 ***	0.59	2.86 **	χ2 = 2.30, p = 0.13
Support → Project Performance	0.35	3.25 ***	0.65	3.45 ***	χ2 = 3.19, p = 0.07

Table 10: Comparison between US and Thailand without moderation effect

REFERENCES

Andersen, E. S., et al. (2006). Exploring project success. *Baltic Journal of Management*, *1(2),* 127-147.

Anderson, J.C. and Gerbing, D.W. (1984). “The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory analysis”, *Psychometrika*, 49, 155-173.

Atkinson, R. (1999), Project management: cost, time and quality, two best guesses and a phenomenon, it´s time to accept other, *International Journal of Project Management*. 17(6), 337-342.

Aubry, M. (2015). Project management office transformations; Direct and moderating effects that enhance performance and maturity. *Project Management Journal*, 46(5), 19-45.

Bagozzi, R.P. (1980) *A Causal Model in Marketing,* New York, Wiley.

Besteiro, E. N., de Souza Pinto, J., & Novaski, O. (2015). Success factors in project management. *Business Management Dynamics, 4(*9), 19-34*.*

Bonner, J.M., Ruekert, R.W., Walker Jr., O.C., 2002. Upper management control on new product development projects and process performance. Journal of Product Innovation Management 19, 233–245.

Braunscheidel, M. and Suresh, N., 2009. The organizational antecedents of a firm’s supply chain agility for risk mitigation and response. *Journal of Operations Management,* 27(2), 119-140.

Brown, S.L., & Eisenhardt, K.M. (1995). Product development: past research, present findings and future directions. *The Academy of Management Review 20*(2), 343–378.

Browne, M.W. and Cudeck, R. (1993). “Alternative ways of assessing model fit”. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models.* (pp. 445-455). Newbury Park, CA: Sage.

Carvalho, M.M., Patah, L.A., and Bido, D.S., (2015). "Project management and its effects on project success: Cross-country and cross-industry comparisons", *International Journal of Project Management*, 33, pp.1509-1522.

Cao Hao, T., & Swierczek, F. W. (2010). Critical success factors in project management: implication from Vietnam. *Asia Pacific Business Review*, *16*(4), 567-589.

Chou, J. S., Yang, J. G., 2012. Project management knowledge and effects on construction project outcomes: an empirical study. *Project Management Journal*. 43(5), 47-67.

Churchill Jr., G.A., (1979). "A paradigm for developing better measures of marketing constructs". Journal of Marketing Research, 16 (February), 64–73.

Cordoba, J., & Piki, A. (2011). Facilitating project management education through groups as systems. *International Journal of Project Management, 30*, 83-93.

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology, 71*, 500-507.

Fornell, C., 1982. A second generation of multivariate analysis: An overview. In: C. Fornell, ed., *A second generation of multivariate analysis.* New York, NY: Praeger, 1–21.

Fortune, J., & White, D. (2006). Framing of project critical success factors by a systems model. *International Journal of Project Management*, *24*, 53-65.

Fortune, J., White, D., Jugdev, K., & Walker, D. (2013). An exploratory study of project success with tools, software and methods. *International Journal of Managing Projects in Business, 6(3),* 534-551.

Frame, J.D., 2002. *The new project management: Tool for an age of rapid change, complexity, and other business realities*, Second edition, Jossey-Bass: a Wiley Brand, Indianapolis, IN.

Gray, R. J. (2001). Organizational climate and project success. *International Journal of Project Management, 19*(2), 103-109.

Gu, V. C., Hoffman, J. J., Cao, Q., & Schniederjans, M. J. (2014). The effects of organizational culture and environmental pressures on IT project performance: A moderation perspective. *International Journal of Project Management, 32*(7), 1170-1181.

Guzzo, R.A., and Dickson, N. W. (1996). "Teams in organization: Recent research on performance and effectiveness", *Annual Review of Psychology*, 47, 307-338.

Hackman, J. R. (1990). *Groups that work (and those that don't).* San Francisco, CA: Jossey-Bass Publishers.

Hair, J.F., Hult G.T.M., Ringle, C.M., Sarstedt, M. (2013). *A Primer on Partial Least Squared Structural Equation Modelling,* Los Angeles, CA: Sage Publications.

Hair, J., Black, W., Babin, B., Anderson, R., and Tatham, R. (2006), Multivariate Data Analysis, Sixth edition, Pearson Prentice Hall, Upper Saddle River, New Jersey.

Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (1998). *Multivariate Data Analysis*. Fifth Edition, Macmillan Publishing Company, New York, NY.

Henderson, J. C., & Lee, S. (1992). Managing I/S design teams: A control theories perspective. *Management Science, 38*(6)*,* 757-777*.*

Henseler, J., Ringle, C. and Sinkovics, R., 2009. The use of partial least squares path modeling in international marketing. *New Challenges to International Marketing,* 20*,* 277-319.

Hirst, G. (1999). *The relationship between communication and R&D project performance: A five factor model of team communication.* University of Melbourne, Australia.

Hsu, S.H., Chen, W.H. and Hsieh, M.J., 2006. Robustness testing of PLS, LISREL, EQS, and ANN-based SEM for measuring customer satisfaction. *Total Quality Management,* 17(3), 355-371.

Hu, L. and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria version new alternatives. Structural Equation Modeling, 6(1), 1-55.

Hyväri. I. (2006). Success of projects in different organizational conditions. *Project Management Journal.* International project management standard IS021500. Australian Committee MB12. Disponível em: http://www.mb12.org.au/.

Igbaria, M., and Tan, M. (1997). The consequences of information technology acceptance on subsequent individual performance, *Information & Management*, 32, 113-121.

Jahanshahi, A. A., and Brem, A. (2017). Does real options reasoning support or oppose project performance? Empirical evidence from electronic commerce projects, *Project Management Journal*, August/September, 39-54.

James, L.R., Mulaik, S.A. and Brett, J.M., 1982. Causal Analysis: Assumptions, Models and Dada, Sage Publications. Beverly Hills, CA.

Jiang, J. J., Klein, G., & Means, T. L. (2000). Project risk impact on software development team performance. *Project Management Journal, 31(4)*, 19-26*.*

Keil, M., Bernard C. Y. Tan, Wei, K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. *MIS Quarterly, 24*(2), 299-325.

Kemery, E. R. & Stickney, L. T. (2014). A multifaceted approach to teamwork assessment in an undergraduate business program. *Journal of Management Education, 38*(3), 462-479.

Kline, R.B. (2010). *Principles and Practices of Structural Equation Modelling,* Third Ed., New York: The Guilford Press.

Langley, M. A. (2015) Capturing the value of project management: Through knowledge transfer. *Project Management Institute, Inc*. pmi/org/pulse.

Larson, E. & Drexler, J. A. (2010). Project management in real time: A service-learning project. *Journal of Management Education, 34*(4), 551-573.

Lee, D., 1997. The impact of poor performance on risk-taking attitudes: a longitudinal study with a PLS causal modeling approach. *Decision Sciences,* 28(1), 59-80.

Leek, J.T., & Peng, R.D. (2015). What is the question? *Science, 347,* 1314-1315.

Leung, H. (2001). Organizational factors for successful management of software development. *Journal of Computer Information Systems, 42(2)*, 26-37.

Liu, S., and Deng, Z. (2015). How environment risks moderate the effect of control on performance in information technology projects: Perspectives of project managers and user liaisons. International Journal of Information Management, 35(1), 80-97.

Lundin R .A. & Soderholm A. (1995). A theory of the temporary organization. *Scand. J. Management. 11(4),* 437-455.

MacCallum, R.C., Browne, M.W., and Sugawara, H.M. (1996). “Power analysis and determination of sample size for covariance structure modeling.” Psychological Methods, 1, 130-149.

Martin, N. L., Pearson, J. M., & Furumo, K. (2007). IS project management: Size, practices, and the project management office. *Journal of Computer Information Systems, Summer,* 52-60.

Mathur, G., Jugdev, K., & Shing Fung, T. (2014). The relationship between project management process characteristics and performance outcomes. *Management Research Review, 37*(11), 990-1015.

McFarlan, F. W. (1981). Portfolio approach to information systems. *Harvard Business Review, 59(5),* 142-150.

Meredith, J. R., Mantel, S. J., & Shafer, S. M. (2015). *Project Management: A Managerial Approach* (9th ed.), New York: John Wiley & Sons.

Mir, F. & Pinnington, A. (2014). Exploring the value of project management: Linking project management performance and project success. *International Journal of Project Management, 32(2),* 202-217.

Napier, N., Kel, M., & Tan, F. (2007). IT project managers’ construction of successful project management practice: A repertory grid investigation. *Information Systems Journal, 19*, 255-282.

Nunnally, J.C. (1978). Psychometric Theory, McGraw Hill, New York.

Ofori, D. F. (2013). Project management practices and critical success factors: A developing country perspective. *International Journal of Business and Management, 8*(21), 14-31.

Ozpolat, K., Chen, Y., Hales, D., Yu, D., & Yalcin, G.M. (2014). Using Contests to Provide Business Students Project-Based Learning in Humanitarian Logistics: PSAid Example, *Decision Sciences Journal of Innovative Education,* 12(4), 269-285.

Pinto, J.K., Kharbanda, O.P., 1996. *Successful Project Managers: Leading Your Team to Success*. Van Nostrand, New York.

Pinto, J.K., & Slevin, D.P. (1987a). Critical factors in successful project implementation. *IEEE Transactions of Engineering Management, 34*(1), 22–27.

Pinto, J.K., & Slevin, D.P. (1987b). Balancing strategy and tactics in project implementation. *Sloan Management Review 29*(1), 33–42.

Porter, W. T., and Lilly, S. B. (1996). The effects of conflict, trust, and task commitment on project team performance, *The international Journal of Conflict Management*, 7(4), 361-376.

Project Management Institute (2021). *A Guide to the Project management Body of Knowledge* (PMBOK® Guide) 7th ed., Project Management Institute, Newtown Square, PA.

Raiden, A. B., and Dainty, A. R. J. (2006). Human resource development in construction organisations: An example of a "chaordic" learning organization, *The Learning Organization*, 13(1), 63-79.

Riou, J., Guyon, H., & Falissard, B. (2016). An introduction to the partial least squared approach to structural equation modelling: a method for exploratory psychiatric research. *International Journal of Methods in Psychiatric Research, 25(3),* 220-231.

Rosenzweig, E.D., 2009. A contingent view of e-collaboration and performance in manufacturing. *Journal of Operations Management,* 27(6), 462-478.

Sambamurthy, V. and Chin, W.W., 1994. The effects of group attitudes toward alternative GDSS designs on the decision-making performance of computer-supported groups. *Decision Sciences,* 25(2), 215-241.

Saucer, C., & Horner-Reich, B. (2009). Rethinking IT project management: Evidence of a new mindset and its implications. *International Journal of Project Management, 27*, 182-193.

Scott-Young, C. & Samson, D. (2008). Project success and project team management: Evidence from capital projects in the process industries. *Journal of Operations Management, (26)*, 749-766.

Settle-Murphy, N., & Thornton, C. (1999). Facilitating your way to project success. *Information Strategy: The Executive's Journal, 15(3),* 6-44.

Shenhar, A.J. (2001), "One size does not fit all projects: exploring classical contingency domains", *Management Science*, 47(3), pp. 394-414.

Soderlund, J. (2002). Managing complex development projects: arenas, knowledge processes and time. *R&D Management, 32(5),* 419-430.

Swafford, P.M., Ghosh, S. and Murthy, N., 2006. The antecedents of supply chain agility of a firm: scale development and model testing. *Journal of Operations Management,* 24(2), 170-188.

Tabassi, A. A., Roufechaei, M. K., Baker, A. H. A. and Yusof, N. (2017). "Linking team condition and team performance: a transformational leadership approach", *Project Management Journal,* April/May, 22-38.

Thamhain, H., (1990). Managing technologically innovative team efforts toward new product success. *Journal of Product Innovation Management 7(1),* 5-18.

Thamhain, H., & Wilemon, D. (1987). Building high performing engineering project teams. *IEEE Transactions of Engineering Management 34*(3), 130–137.

Verburg, R. M., Bosch-Sijitsema, P., & Vartiainen, M. (2013). Getting it done: Critical success factors for project managers in virtual work setting. *International Journal of Project Management, 31*(1), 68-79.

Yazici, H. J. (2009). The role of project management maturity and organizational culture in perceived performance. *Project Management Journal, 40*(3), 14-33.

Yetton, P., Martin, A., Sharma, R., & Johnston, R. (2000). A model of information systems development project performance. *Information Systems Journal, 10(4),* 263-289.

Wysocki, R. K., Beck, Jnr. R., & Crane, D. B. (2000). Effective Project Management (2nd ed.). New York: John Wiley & Sons, Inc.

Zmud, R. W. (1980). Management of large software development efforts. *MIS Quarterly, 4(2),* 45-55.

DECISION SCIENCES INSTITUTE

How Supplier Diversity and Enabled Minority Businesses Create Stakeholder Value in Underserved Communities

Ashok Vairavan
Georgia State University
avairavan1@gsu.edu

G. Peter Zhang
Georgia State University
gpzhang@gsu.edu

ABSTRACT

While large purchasing organizations (LPOs) implement supplier diversity (SD) programs to support minority business enterprises (MBEs) and create value-added diverse supply chains, many SD initiatives fall short of enabling MBEs and generating the desired impact. This paper explores how supplier diversity can be transformative in helping grow and enable MBEs to create stakeholder values in underserved communities. This study takes an MBE focused approach to examine the efficacy of SD programs, the relationship facets that underpin mutually beneficial relationships between LPOs and MBEs, and the enabled MBEs' impact on their stakeholders' economic and social well-being.

KEYWORDS: Minority business enterprises, Stakeholders, Supplier diversity, Supply chain management

INTRODUCTION

The Business Roundtable, a leading business association comprised of CEOs of the nation's largest companies, issued a statement to serve all stakeholders as the new purpose of corporations (Business Roundtable, 2019), countering the long-held neo-economic view that corporations' sole purpose is to generate returns for their shareholders (Friedman, 2001). One of the initiatives that help corporations to achieve this renewed purpose of serving various stakeholders and gaining increasing prominence is supplier diversity (SD), in which large purchasing organizations (LPOs) source products and services from diverse suppliers such as minority business enterprises (MBEs). The National Minority Supplier Development Council (NMSDC) defines an MBE as a firm that is at least 51 percent owned and operated by a minority in one of the racial categories of African, Asian-Pacific, Hispanic, and Native American.

In addition to the increased corporate focus on pursuing a stakeholder-centric strategy (Schwab, 2021) in which companies seek long-term value creation by meeting the needs of all their stakeholders, several recent events and trends have driven the renaissance of supplier diversity programs, which are aimed at supporting MBEs and broader stakeholder groups, such as employees and communities. These include the disproportionate effects of the Covid-19 pandemic on minority-owned businesses (Dua et al., 2020), the escalated racial tensions and social unrest, the ongoing global supply chain disruptions caused by the pandemic, trade protectionism, and geopolitical tensions and the population trends in which minorities are projected to constitute over 50 percent of the expected total population by 2044 (Colby & Ortman, 2015). A strategic avenue for LPOs to expand their customer base and gain valuable market insights is to diversify their supply chains by partnering with MBEs (Adobor & McMullen, 2007; Worthington, 2009).

This study aims to expand SD research by exploring the effectiveness of SD mainly from the perspective of MBEs and their stakeholders. We ask the following questions: (1) how do MBEs perceive the efficacy of LPO's supplier diversity programs? (2) how can MBEs and LPOs cultivate reciprocal relationships, enabling MBEs to grow to scale? and (3) how do enabled MBEs generate a well-being impact for their stakeholders, catalyzing a virtuous cycle in the LPO-MBE-stakeholder relationship? To answer these questions, we adopt a multi-case study design drawing on semi-structured interviews with informants from MBEs, LPOs, and various stakeholder groups. We use social exchange theory and stakeholder theory as our theoretical lenses to examine how LPOs and MBEs can impact a broad set of stakeholders beyond the traditional buyer-supplier relationships and for predicting positive outcomes in interfirm relationships.

LITERATURE REVIEW

Prior research has examined both demand- and supply-side factors that motivate LPOs to implement SD initiatives (Richard et al., 2015; Slater et al., 2008; Worthington, 2009). These factors include attracting new diverse customers, mitigating supply chain risks, and improving innovation and product quality (Blount, 2021; Greer et al., 2006; Shah & Ram, 2006). On the other hand, as minorities represent an increasingly larger proportion of the population and comprise a growing sector of the economy, minority businesses can be a source of innovation (Porter, 2019). In addition, because of their relatively small size, MBEs are more agile and flexible in responding to changing external environments and customer demand, thus reducing supply chain risks (Shah & Ram, 2006).

For MBEs, the complexity of doing business with large firms often represents the most significant hurdle to developing healthy interfirm relationships (Dollinger et al., 1991). Greer et al. (2006) find that factors that hinder the dyadic relationship effectiveness include bilateral cultural and communication gaps between LPOs and MBEs, MBEs' financial challenges, and their lack of understanding of corporate politics navigating the LPO bureaucracy. They show that LPOs' human resource diversity practitioners could use their expertise to help close cultural gaps between LPOs and MBEs and promote the business case within the LPO organizations. Another inhibiting factor to a healthy interfirm relationship is the gap between LPOs' espoused rhetoric to supplier diversity and their actions in supporting, developing, and growing MBE suppliers (Blount, 2021). This is often caused by the asymmetrical power in the relationship between two firms of unequal size. Using a fixed-pie, zero-sum approach and a predisposition to follow the traditional corporate procurement playbook, LPOs are likely to exert maximum bargaining power on MBEs, who have little choice but to accept onerous contract terms (i.e., pricing and payment terms), which can severely constrain MBEs' ability to reinvest in their business (Lashley & Pollock, 2020; Richard et al., 2015). While this practice may help LPOs with short-term benefits, it would negatively affect their supply chain's overall performance by constraining MBEs' ability to develop resources and capabilities that enhance their products and services (Greenhalgh & Lowry, 2011), lowering MBEs' trust and commitment to the relationship (Maloni & Benton, 2000). Therefore, if MBEs are not treated fairly, it would not be in the best interest of LPOs in the long run (Adobor & McMullen, 2007; Griffith et al., 2006).

Previous studies have situated research issues at the buyer or dyadic relationship level and have paid less attention to how SD impacts MBEs' stakeholders beyond LPOs. Because the buyer-supplier relationship is embedded in a complex network of stakeholder relationships, Ulaga and Eggert (2006) suggest that research should examine value creation in the broader environment surrounding buyers and their key suppliers. Addressing social issues within a sustainable supply chain management framework, Bals and Tate (2018) find that the existing research predominantly focuses on direct and economic stakeholders in supply chains and the economic and environmental dimensions but generally overlooks secondary stakeholders and the supply chain's impact on social welfare. Lee (2021) coins the term "supply chains with a conscience" to emphasize the need for supply chains to maximize the well-being of society. He encourages supply chain scholars to expand supply chain research by including broader impact measures such as the well-being of people, communities, and social enterprises. Sodi et al. (2022) point out the limited well-being impact of traditional SD programs on economically disadvantaged populations and call for a move from supplier diversity to economic inclusion to have a broad community impact.

THEORETICAL FOUNDATION

Social exchange theory and stakeholder theory provide the theoretical context for our study. Specifically, we draw on social exchange theory (Blau, 1964) to frame the dyadic relationship between an LPO and an MBE and stakeholder theory (Freeman, 2001) as the overarching framework to examine the interdependent relationships between MBEs and their stakeholders. Social exchange theory has been widely applied in supply chain research to understand interfirm relationships (e.g., Griffith et al., 2006). An underlying premise of social exchange theory is the significance of trust and commitment in mutually beneficial relationship formation (Ambrose et al., 2010). Moorman et al. (1992) define trust as the willingness to rely on an exchange partner in whom one has confidence. Trust between parties can be generated in two ways: (1) through frequent, reliable, and consistent reciprocation of benefits, and (2) through the gradual expansion of exchanges (Blau, 1964). One of the theory's main thrusts is that powerful actors in an exchange relationship are better off downplaying their power and concentrating instead on building a mutually beneficial relationship (Emerson, 1976). Exchange partners' satisfaction levels become the prime determinants of whether future exchanges will occur or not (Miles, 2012). This dynamic is apropos for the LPO-MBE relationship, where the MBE is often the weaker firm within the dyad.

The central premise of stakeholder theory is that organizations should focus on meeting a broader set of interests than just amassing shareholder wealth (Freeman, 2001). The theory expands upon the traditional view of firms, which argues that a firm's primary objective is to maximize wealth for its shareholders, irrespective of the direct value accruing to other stakeholders. In the shareholder model, firms base their decisions primarily on economic terms for shareholders without explicitly considering other stakeholder interests. Freeman et al. (2004) take an instrumental view of stakeholder management by arguing that while shareholders are essential stakeholders, firm profits are the result rather than the driver of value creation for all other stakeholders. In other words, effectively managing a diverse set of stakeholders will lead to profits, rather than the counterview that prioritizing shareholder value will lead to stakeholder benefits.

We use the stakeholder theory as an encompassing framework to examine our research questions, with MBE firms as the primary focus. The internal stakeholders of an MBE include frontline and managerial employees and board members. External stakeholders consist of customers (LPOs), government officials, suppliers, and communities. Stakeholder theory can also be framed from the LPO perspective in that LPOs may aim to diversify their supply chains as part of their corporate social responsibility initiatives to support MBEs and their communities. However, LPOs may confront a tension between managing stakeholder and shareholder interests by balancing their longer-term stakeholder goals and obligations with their short-term priorities of increasing profits and supply chain efficiencies by minimizing their transaction costs and exerting bargaining power with their suppliers (Adobor & McMullen, 2007; Greenhalgh & Lowry, 2011).

METHODS

Because our research aims to explore how LPOs and MBEs can work together through SD to enable MBEs to grow and create broad stakeholder impact and to develop a theoretical framework linking SD, MBE-LPO relationship, and stakeholder wellbeing, the case study is an appropriate research method (Yin, 2018). In choosing our cases, we are guided by the purposeful replication logic proposed by Yin (2018).

Case Selection

Leveraging our professional network and industry leaders' referrals, we recruited and selected five MBE supplier firms with the following characteristics: (1) certified by the National Minority Supplier Development Council, (2) in business for a minimum of five years, (3) operating in different industries, and (4) a minimum of five years of experience supporting LPOs with SD programs. We used the Chief Executive officers (CEOs) as our primary MBE informants because of their intimate knowledge of customer relationships and their extensive experience with SD (Zaheer & Venkatraman, 1995). Table 1 provides a summary of the MBEs in our sample.

Table 1. A Summary of MBE Firms

Case Pseudonym	Firm Overview	Research Questions
MBE1 (Focal Firm)	• Industry: Business Process Outsourcing • Solutions: Customer Contact Center Solutions • Years in Business: 6 • Total Employees: >2,000 (95% minority)	1, 2, 3
MBE2	• Industry: Information technology • Solutions: software development • Years in Business: 10 • Total Employees: 25 (90% minority)	1, 2
MBE3	• Industry: Supply chain logistics and packaging • Solutions: Packaging solutions • Years in Business: 7 • Total Employees: 250 (55% minority)	1, 2

MBE4	• Industry: Technology Services and Staffing • Solutions: global digital IT staffing and consulting services • Years in Business: 13 • Total Employees: 890 (52% minority)	1, 2
MBE5	• Industry: Food industry • Solutions: custom seasoning blends, food & beverage mixes • Years in Business: 15 • Total Employees: 75 (90% minority)	1, 2, 3

In addition to the five MBE firms, we interviewed SD leaders from four LPO firms. We include LPOs in the study to gain a balanced perspective and a more holistic understanding of the buyer-seller relationship (Graebner & Eisenhardt, 2004). Table 2 provides a summary of the LPOs in our sample.

Table 2. Summary of LPO Informants

	LPO1	LPO2	LPO3	LPO4
LPO Industry	Healthcare	Insurance	Telecommunications	Financial Services
Length of Relationship with MBE1	6 years	4 Years	2 Years	6 Years
Informant Title	Manager for Impact spending	Director of Supplier Inclusion	Director of Supplier Diversity	Diversity Solutions Leader

To address our third research question, we interviewed internal and external stakeholders associated with MBE1 and MBE5. Table 3 summarizes the stakeholders we interviewed.

Table 3. Summary of Stakeholder Informants

MBE	Stakeholder Category	Title	Location
MBE1	Company Leader	CEO	Morrow, GA
MBE1	Employee	Customer Service Representative	Morrow, GA
MBE1	Manager	Quality Manager	Morrow, GA
MBE1	Employee	Customer Service Representative	Morrow, GA
MBE1	Manager	Team Manager	Morrow, GA
MBE1	Manager	Team Manager	Dallas, TX
MBE1	Employee	Customer Service Representative	Dallas, TX
MBE1	Employee	Customer Service Representative	Dallas, TX
MBE1	Manager	Team Manager	Dallas, TX
MBE1	Supplier	CEO - Supplier	Morrow, GA
MBE1	Supplier	COO - Supplier	Dallas, TX
MBE1	Community	President Minority Supplier Development	Georgia
MBE1	Government	Chairman Clayton County Board	Morrow, GA
MBE1	Government	Councilmember 1	Dallas, TX
MBE1	Community Developer	Managing Partner	Dallas, TX
MBE1	Board Member	Board Member / Investor	New York
MBE5	Company Leader	CEO	Colombus, OH
MBE5	Employee	Operations Analyst	Colombus, OH
MBE5	Supplier	Account Executive	Colombus, OH
MBE5	Community Member	Educator/Community Leader	Colombus, OH

Data Collection

Our primary data source is semi-structured interviews collected from informants varied across three separate groups: MBEs, LPOs, and stakeholders. We also used supplemental data sources consisting of investor diligence documents, electronic exchanges between MBEs and LPOs, observations in MBE-LPO meetings, corporate SD websites, a Harvard Business School Teaching Case written on the focal firm, supplier contracts, and practitioner webinars, enhancing the findings' reliability and validity (Yin, 2018).

For our primary source of data, we developed separate interview guides for MBEs and LPOs to understand the different perspectives of suppliers and buyers and a third interview guide specifically for stakeholders to understand the well-being impact across different groups associated with enabled MBEs. Each interview conducted with an MBE informant lasted between 90-120 minutes. The interviews with the four SD leaders lasted between 60 to 90 minutes and focused on the background and goals of their programs, challenges, and barriers to doing business with MBEs, as well as their views on how MBEs impact their stakeholders. Stakeholder interviews averaged between 45 to 60 minutes, determined by the stakeholder category (i.e., employees, suppliers, community leaders, and investors), and focused on the MBE's impact.

Coding and Analysis

We developed three provisional coding schemes for each informant group (e.g., MBE Suppliers, LPO Buyers, and Stakeholders) a priori based on the interview questions' content, the literature, and the core tenets of social exchange and stakeholder theories (Miles & Huberman, 1994). As the interviews progressed and secondary data sources were analyzed, we created additional codes based on the emergent themes derived from the data. Supplemental observational data were included in our case database as analytic memos and coded separately. In this process, we combined deductive and inductive coding to develop a complete set of first-order codes.

In total, we generated 35 first-order codes for MBEs, 31 first-order codes for LPOs, and 22 first-order codes for stakeholders. This second coding cycle was conceptional and focused on the interrelationships of categories constructed to develop higher-level analytic concepts (Miles et al., 1994). The second-level analysis yielded nine second-order codes for MBEs, eight second-order codes for LPOs, and five second-order codes for stakeholders. Themes were subsequently analyzed across cases to establish patterns and formulate theoretical constructs, ensuring that emerging constructs were abstracted from the context (Eisenhardt, 1989).

RESULTS

MBE Perspective on SD

MBEs noted that their companies' growth was primarily driven by the differentiated solutions that provide strategic value to their customer and the quality of the relationships they forged with their customers. Having the MBE-certified status alone is not sufficient to win new business. However, the designation helps MBEs gain exposure to potential supplier opportunities through LPOs' SD programs.

Despite the positive movement in the corporate community, the MBE's perceived effectiveness of SD programs in helping MBEs grow is low. Specifically, MBEs believe that many LPOs lack the company-wide commitment to achieving their programs' espoused goals, which creates a sentiment of indifference and complacency. A sentiment that emerged consistently from these MBEs is the belief that many LPOs “check the box” in approaching the relationship with MBEs. In addition, LPO SD initiatives often do not have sufficient resources to effectively implement and achieve their mission and goals, which leads to questions about SD’s legitimacy within the LPO. MBE5 perceived many SD programs as “toothless sharks” within their companies, a sentiment shared by the other MBE firms. This perspective is based on their experiences that SD professionals have limited influence and do not have the authority for supplier buying decisions with their firms. Amplifying this belief is the fact that while there may be hundreds of product categories LPOs source from suppliers, their SD programs often do not have sufficient resources or manpower to identify, support, and add MBE suppliers.

For SD to be more effective and impactful, MBEs offered several suggestions. First, LPOs must signal strong commitment through top management support and sufficient resource allocation for SD efforts. Second, defining SD goals and objectives that ladder up to the strategic corporate goals is key for SD to gain legitimacy and attention throughout the LPO firm. Third, each MBE CEO noted that while LPOs often publicly promote their procurement spend with minority businesses, they need to change their approach with how they negotiate with MBEs to ensure that MBEs can thrive in the relationship.

Relationship Building

MBEs and LPOs engage in supplier partnerships with the intention that the relationship will result in mutually beneficial outcomes such as improved competitiveness, stronger financial performance, and a positive impact on stakeholders. The following relationship qualities emerged from the data contributing to mutually beneficial outcomes for LPOs and MBEs: differentiated solutions that bring strategic value, aligned shared values, and high-quality interfirm communications. These factors are salient antecedents to trust and commitment developed between MBEs and LPOs.

Differentiated Strategic Solutions

The MBEs in our study emphasized the importance of delivering differentiated, strategic solutions to LPOs, which significantly influences whether a relationship will flourish in the long run. Strategic products and services offered by suppliers add long-term economic value to the buyers’ value chain and are a source of gaining and sustaining a competitive advantage. Following a differentiated product strategy, MBEs can create higher value for customers, thereby allowing them to charge premium prices and become more entrenched within the LPOs’ supply chains. This strategy is contrasted with a cost-leadership strategy, where firms work to optimize their value chain activities and offer standardized products and services to achieve a low-cost position in its industry that many MBEs follow by producing commodity solutions, fueling the LPO perception that MBEs offer low-value and substitutable products and services.

Shared Values and Interfirm Communications

The views from MBEs and LPOs converge on the significance of having compatible operating norms, shared values, and a long-term orientation for building successful buyer-supplier relationships. We define shared values as the fit or congruence of firms with inclusive and collaborative cultures (McAfee et al., 2002). Cooperation and inclusivity must be engrained in the LPO’s culture, which requires top management commitment. LPO employees need to understand that it is in the firm's best interests to do business with MBEs and that SD is not just a corporate social responsibility program (Adobor & McMullen, 2007).

As a relational competency, effective communication enables supply chain partners to interact and exchange tacit information, thus leading firms to identify market opportunities and solve problems together (Ulaga & Eggert, 2006). Joint value chain activities and frequent interactions between LPO and MBE provide opportunities for strengthening communication, engendering trust, and creating closer, collaborative relationships. Effective interfirm communication requires that firms adopt a collaborative approach to building strategic capabilities that benefit both parties. Therefore, if an LPO approaches the relationship with the view that an MBE is a strategic partner rather than a transactional vendor, it will prioritize richer dialogue and collaboration with a long-term orientation.

Power Moderator

Power asymmetry in an LPO-MBE relationship is a function of firm size, the percentage of revenue that the LPO contributes to the MBE, the value associated with the product or service, and perceived equity (Harland et al., 2004). MBEs noted the difficulties they have experienced when LPOs exert their power in negotiating and dealing with MBEs. Benton and Maloni (2005) found that exploitation of the supply chain by a powerful partner may lead to resentment and underperformance, thus hurting the power holder, and that judicious use of power may serve to benefit the power holder. If an LPO exerts its maximum bargaining power upon MBEs, it can have a harmful cascading effect on the MBEs' key stakeholders, namely their employees, particularly in labor-intensive industries. MBEs would struggle to attract new employees with the necessary skill levels and engender satisfaction and commitment from existing employees. Our findings on power asymmetry in the LPO-MBE relationship support the main construct in the social exchange theory that powerful actors in an exchange relationship are better off downplaying their power.

Impact on Stakeholders

The LPO-MBE partnership can have a far-reaching impact beyond the dyadic relationship as MBEs rely on the stakeholder ecosystem to support their value-chain activities and improving the relationship with various stakeholders whose support has been linked to above-average returns (Martinez et al., 2017) can help MBEs become more valuable to LPOs, fueling their continued growth.

Impact on Employees

Employees constitute one of the most critical resources and stakeholder groups that help companies gain and sustain competitive advantages (Barney, 1991). MBE1's employee demographics reflect that of the underserved communities in which it operates, with 93% of its employees African American and most living within a 10-mile radius of their call centers. MBE1's Morrow, GA call center employs approximately 1,000 full-time employees, with 90% of its total workforce as front-line customer service representatives tasked with providing various customer contact services on behalf of its customers. MBE1's call center in Dallas has 500 full-time employees with similar employee demographics and offers similar customer care services as its Morrow location. Employees are recruited and onboarded to support clients through a comprehensive client-specific training program to gain technical and soft skills and domain knowledge specific to the client and the client's industry.

MBE1's business model is driven by its employee-centered culture. It offers employees competitive living wages, comprehensive health and welfare benefits, and other professional development programs. The company has made a conscious decision to pay its front-line employees wages up to 30% higher than the industry standard due to its ability to charge premium prices to customers based on its differentiated business model. Also, MBE1 is committed to promoting employees into leadership roles, where 80% of its managers started as customer service agents, which is made possible by the growth they have experienced with LPO customers.

Analysis of employee data from MBE5 yielded similar results. While a smaller firm operating in a different industry than MBE1, MBE5 is similarly located in an underserved community, with its workforce predominantly minority. MBE5 offers professional training to all employees ranging from job-related operational tasks to more complex and transferrable skillsets such as lean six sigma, decision-making, and leadership through partnerships with its clients and local universities. Further, as MBE5 is transformed into a more strategic supplier, its employees' wages and benefits have increased significantly.

Impact on Suppliers

MBE1 and MBE5's growth and expansion required suppliers to support their operations in technology, security, raw materials supply, facilities management, and human capital. Because most of their suppliers are also MBEs and located in the same communities, their growing businesses with LPOs have had a positive cascading effect on these suppliers. The supplier firms we interviewed were minority-owned, including a security services firm, a facilities management solutions provider for MBE1, and a specialty ingredients provider for MBE5. These suppliers have forged long-term contracts with the MBEs with products and services central to MBEs' core operations. The suppliers noted that not only did their companies experience growth, but there was broad community revitalization resulting from the MBEs' presence in the communities.

With the attention the enabled MBEs garnered in the community, their suppliers appeared to ride on the "coattails" of MBEs' respective success and reputation. Because the suppliers were local firms that were small and relatively unknown, partnering with enabled MBEs allowed them to grow their business and enhance their brand image in the community. As enabled MBEs continue to grow, they will rely more on suppliers to support their operational needs, suggesting another positive feedback loop where enabled MBEs impact their suppliers by helping them grow while drawing on suppliers' resources and capabilities to support their operations.

Impact on Investors

Access to capital plays an essential role for MBEs to grow. Yet, it remains one of the biggest challenges that all MBEs face. During our research, MBE1 was recapitalizing its business, resulting in the company closing an investment with a private equity firm, which provided growth capital and credibility when dealing with larger LPOs. The private equity firm secured two board seats, however, the updated capital structure allowed MBE1 to retain its certified minority designation as the majority equity holder. As part of a rigorous due diligence process, the investment firm concluded that MBE1's minority status and social mission aligned well with macro-diversity trends in business today as socially conscious companies prioritized their SD programs to promote an inclusive approach to procurement.

Like MBE1, MBE5 developed reciprocal relationships with its financiers, allowing the firm to invest in the resources and develop capabilities transforming itself from a commodity supplier to a partner offering differentiated and strategic solutions. MBE5 relied primarily on early angel investors and a bank credit line to finance its growth. Early equity investors were rewarded with positive returns when MBE5 recently sold its business to a larger MBE due to its successful growth and expansion with LPO partners.

Impact on Communities

Community organizations and municipalities are key external stakeholders that can affect and be affected by MBEs who establish their operations in these locations. When MBE1 announced its decision to open contact centers in the two communities, there was considerable enthusiasm for the jobs created and the potential economic impact it would have on local businesses. This enthusiasm is reflected in the following comment by the real estate development firm's CEO that supported MBE1 expansion in Dallas, TX:

> *"The day they had a job fair in the center court and [MBE1] hired 350 people at a living wage rate was one of the most exciting, joyful days I've ever had at work. Because first of all, there were 1,000 people there, so you knew there was a need, and the people who left felt like they had won the lottery. It shows you the level of need." (CEO of Real Estate Development Company*

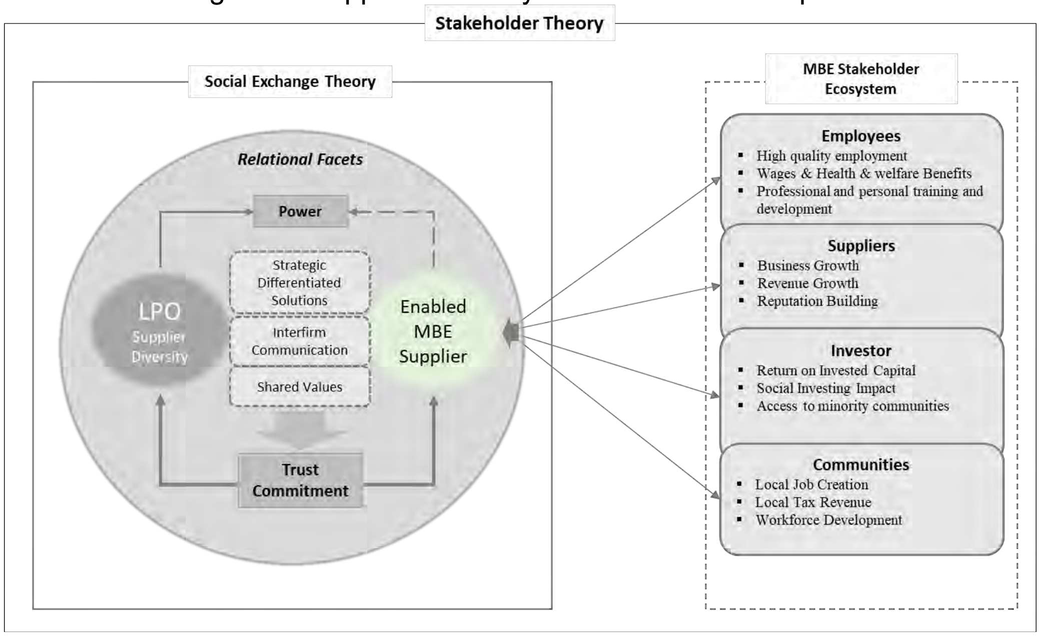
Similarly, MBE5 created a positive impact in the Columbus, OH community. MBE5 also invested significant resources with several organizations in the community, including public schools. For instance, it has worked with local school districts inviting their students to tour its facilities to understand the science behind the food it developed. The company also supported a volunteer program where its employees served as math tutors for underperforming students at local schools.

Our findings show that enabled MBEs can create jobs, pay taxes, and reenergize the communities, which catalyzes a multiplier effect that induces more businesses to open and more job opportunities, higher wages, and community stability. It also keeps economic activity local and provides a better alternative for people to work in the community rather than commuting long distances to larger economic hubs. For example, MBE1 estimates that it has created over $250M in economic impact for Morrow since it opened its call centers in 2016. The total impact included the wages and benefits it provided to all employees between 2016 and 2021, along with the local taxes expenditures and the estimated money spent with local businesses. The tax revenue generated from the MBE's business activities (i.e., payroll taxes and corporate taxes) helped communities fund public services such as public safety, infrastructure improvements, and schools.

<u>Integrated Stakeholder Framework</u>

Analyzing our findings through the lenses of social exchange theory and stakeholder theory, we develop a conceptual framework in Figure 1, showing how enabled MBEs, through their relationships with LPOs, can have a social and economic impact on stakeholders extending beyond the LPO-MBE dyad. The framework provides an integrative perspective of the interdependent relationships among LPOs, MBEs, and stakeholders, as well as the LPO-MBE relationship facets that require development and active management to enable MBEs to create stakeholder impact.

Figure 1. Supplier Diversity and Stakeholder Impact



DISCUSSION AND CONCLUSION

By exploring supplier diversity's well-being impact on supply chain stakeholders, particularly those in underserved communities, we contribute to the emerging research on the role of ethical supply chain management in generating and improving well-being outcomes within and beyond the supply chains. First, this study provides a more complete perspective of SD's challenges and potential impact by adopting the MBE's perspective. Second, this study introduces the concept of an "enabled" MBE. Previous SD literature has generally viewed MBEs uniformly, irrespective of the difference in size, capabilities, and scope of business. However, generalizing findings for MBEs as a homogeneous group may conceal the unique issues facing different types of MBEs and obscure the true impact of high-performing MBEs. Third, we contribute to social exchange theory and stakeholder theory with the development of a relationship-stakeholder framework that brings into focus how the MBE-LPO relationship impacts various stakeholders and how these stakeholders can, in turn, reinforce and support the MBE's resources and capabilities.

Our results offer several practical insights for MBEs, LPOs, and Policy Makers. For MBEs, our finding that enabled MBEs can scale and make significant contributions to their stakeholders beyond LPOs should motivate other MBEs to embark on a journey to become enabled by reexamining their strategies in dealing with LPOs. Enabled MBEs can serve as role models and provide relationship guidance for many other MBEs seeking to grow and break out of the commodity mindset of competing primarily on price for low-value products and services. To improve SD programs' efficacy, LPOs need to understand and act on MBEs' concerns and challenges, which serves as a starting point for designing and implementing more effective SD programs. Finally, this research provides policy implications for government officials and public policymakers. Local governments can attract and support minority businesses by providing incentives through tax abatements, job credits, and public services. Such incentives offer much-needed support to MBEs and go a long way toward helping communities with economic development and improved living standards. In addition, governments could expand their incentives to motivate LPOs to increase their engagement with MBEs by providing tax credits for reaching certain spending thresholds with minority companies.

This study has limitations, which can be addressed in future studies. First, we focused this research on established or enabled MBEs who have experienced success in growing their business. Therefore, our results, particularly the stakeholder well-being findings, may not extend to those with different types of MBEs. Second, an important question unanswered is how MBEs work with LPOs to overcome the relationship barriers and add value to the LPO-MBE supply chain. Finally, this study provides a foundation for future research examining how SD programs may change due to increasing interest in stakeholder capitalism and the recent focus on social justice and racial equity issues. For example, research could examine if and how MBEs are affected by changes in SD programs due to LPOs' renewed intention to support minority communities.

REFERENCES

Adobor, H., & McMullen, R. (2007). Supplier diversity and supply chain management: A strategic approach. Business Horizons, 50(3), 219-229.

Ambrose, E., Marshall, D., & Lynch, D. (2010). Buyer supplier perspectives on supply chain relationships. International Journal of Operations & Production Management, 30(12), 1269-1290.

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99-120.

Benton, W. C., & Maloni, M. (2005). The influence of power driven buyer/seller relationships on supply chain satisfaction. Journal of Operations Management, 23(1), 1-22.

Blau, P. M. (1964). Justice in social exchange. Sociological Inquiry, 34(2), 193-206.

Blount, I., & Li, M. (2021). How buyers' attitudes toward supplier diversity affect their expenditures with ethnic minority businesses. Journal of Supply Chain Management, 57(3), 3-24.

Blount, I. Y. (2021). How do large purchasing organizations treat their diverse suppliers? Minority business enterprise CEOs' perception of corporate commitment to supplier diversity. Business & Society, 60(7), 1708–1737.

Colby, S. L., & Ortman, J. M. (2015). Projections of the size and composition of the U.S. population: 2014 to 2060: Population estimates and projections. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2015/demo/p25-1143.pdf.

Dollinger, M. J., Enz, C. A., & Daily, C. M. (1991). Purchasing from minority small businesses. International Journal of Purchasing and Materials Management, 27(2), 9-14.

Dua, A., Mahajan, D., Millan, I., & Stewart, S. (2020). COVID-19's effect on minority-owned small businesses in the United States. McKinsey & Company.

Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. Academy of Management Journal, 32(3), 543-576.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. Academy of Management Journal, 50(1), 25-32.

Emerson, R. (1976). Social exchange theory. Annual Review Sociology, 2(1), 335-362.

Freeman, R. E. (1984). Strategic management: A stakeholder approach. Boston: Pitman.

Freeman, R. E. (2001). Stakeholder theory and the modern corporation. In T. Donaldson, and P. Werhane, eds., Ethical Theory in Business: Prentice Hall, Englewood Cliffs, NJ, 247-257.

Freeman, R. E., Wicks, A. C., & Parmar, B. (2004). Stakeholder theory and "the corporate objective revisited". Organization Science, 15(3), 364-369.

Graebner, M. E., & Eisenhardt, K. M. (2004). The seller's side of the story: Acquisition as courtship and governance as syndicate in entrepreneurial firms. Administrative Science Quarterly, 49(3), 366–403.

Greenhalgh, L., & Lowry, J. H. (2011). Minority business success: Refocusing on the American dream. Stanford University Press.

Greer, B. M., Maltbia, T. E., & Scott, C. L. (2006). Supplier diversity: A missing link in human resource development. Human Resource Development Quarterly, 17(3), 325-341.

Griffith, D. A., Harvey, M. G., & Lusch, R. F. (2006). Social exchange in supply chain relationships: The resulting benefits of procedural and distributive justice. Journal of Operations Management, 24(2), 85-98.

Harland, C., Zheng, J., Johnsen, T., & Lamming, R. (2004). A conceptual model for researching the creation and operation of supply networks 1. British Journal of Management, 15(1), 1-21.

Lashley, K., & Pollock, T. G. (2020). Dancing with giants: How small women-and minority-owned firms use soft power to manage asymmetric relationships with larger partners. Organization Science, 31(6), 1313-1335.

Benton, W. C., & Maloni, M. (2005). The influence of power driven buyer/seller relationships on supply chain satisfaction. Journal of Operations Management, 23(1), 1-22.

Maloni, M., & Benton, W. C. (2000). Power influences in the supply chain. Journal of Business Logistics, 21(1), 49-74.

Martinez, A. D., Russell, Z. A., Maher, L. P., Brandon-Lai, S. A., & Ferris, G. R. (2017). The sociopolitical implications of firm reputation: Firm financial reputation× social reputation interaction on firm financial performance. Journal of Leadership & Organizational Studies, 24(1), 55-64.

Mcafee, R. B., Glassman, M., & Honeycutt Jr, E. D. (2002). The effects of culture and human resource management policies on supply chain management strategy. Journal of Business Logistics, 23(1), 1-18.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Sage.

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships between providers and users of market research: The dynamics of trust within and between organizations. Journal of Marketing Research, 29(3), 314–328.

Porter, K. K. (2019). Implementing supplier diversity: Driver of Entrepreneurship. Palgrave MacMillan, Cham, Switzerland.

Richard, O. C., Su, W., Peng, M. W., & Miller, C. D. (2015). Do external diversity practices boost focal firm performance? The case of supplier diversity. International Journal of Human Resource Management, 26(17), 2227-2247.

Schwab, K. (2021). Stakeholder Capitalism: A Global Economy that Works for Progress, People and Planet. Wiley: Joboken, NJ.

Shah, M., & Ram, M. (2006). Supplier diversity and minority business enterprise development: Case study experience of three US multinationals. Supply Chain Management: An International Journal, 11(1), 75-81.

Slater, S. F., Weigand, R. A., & Zwirlein, T. J. (2008). The business case for commitment to diversity. Business Horizons, 51(3), 201-209.

Ulaga, W., & Eggert, A. (2006). Value-based differentiation in business relationships: Gaining and sustaining key supplier status. Journal of Marketing, 70(1), 119-136.

Worthington, I. (2009). Corporate perceptions of the business case for supplier diversity: How socially responsible purchasing can 'pay'. Journal of Business Ethics, 90(1), 47-60.

Worthington, I., Ram, M., Boyal, H., & Shah, M. (2008). Researching the drivers of socially responsible purchasing: A cross-national study of supplier diversity initiatives. Journal of Business Ethics, 79(3), 319-331.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods*. Sage, Log Angeles, CA.

Zaheer, A., & Venkatraman, N. (1995). Relational governance as an interorganizational strategy: An empirical test of the role of trust in economic exchange. Strategic Management Journal, 16(5), 373-392.

DECISION SCIENCES INSTITUTE

Impact of Robotic Dispensing System on Medication Dispensing Error Rate, Patient Wait Time and Patient Satisfaction

Rupesh Agrawal
Northern Kentucky University
agrawalr1@nku.edu

Ali Balapour
Northern Kentucky University
balapoura1@nku.edu

Jyothish Philip
Northern Kentucky University
philipj1@mymail.nku.edu

ABSTRACT

With the growing prevalence of chronic diseases, including cancer, diabetes, cardiovascular disorders, respiratory problems, and renal disorders, the global pharmaceutical market is projected to reach ~$11 billion by 2030 (CAGR 8.3%). However, evidence for adopting a Robotic Dispense System (RDS) remains limited. The current study evaluates the impact of a robotic dispensing system (RDS) in an outpatient pharmacy. Descriptive analysis and structural equation modeling (SEM) results suggest that post-RDS patient wait time reduced from 35.41 to 17.29 minutes, patient satisfaction improved from 68.44% to 84.50%, and the medication error rate reduced from 0.103% to 0.022%. These changes are statistically significant ($p<.0001$).

KEYWORDS: Robotic dispensing system, Waiting time, Satisfaction, and Dispensing error

INTRODUCTION

The current study evaluates the impact of the RDS in an outpatient pharmacy for a UAE-based hospital. The UAE has a diverse and expanding population, estimated to grow to ten and a half million by the end of 2023. Additionally, the growing prevalence of chronic diseases, such as cancer, diabetes, cardiovascular disorders, respiratory problems, and renal disorders, are driving the growth of pharmaceutical markets. In 2022, the size of the pharmaceuticals market was over $4 billion, with a CAGR of 7.96%. The global market size is estimated to reach over $11 billion by 2030, with a CAGR of 8.3% (Pharmacy Automation System Market Overview, 2020-2030, Ramachandram et al., 2022). The robotic supports pharmacists with large prescription orders by automating counting, sorting, and dispensing medication. The COVID pandemic has accelerated the need for RDS. However, evidence needs to be more comprehensive regarding the impact of medication dispensing error rate, patient wait time, and satisfaction to motivate the widespread adoption of RDS. The current study evaluates the data on patient wait time, satisfaction, and dispensing error rate collected over 24 months to assess the impacts of the RDS. Results support the hypothesis that the RDS reduces medication dispensing error rate and patient wait time and improves patient satisfaction.

LITERATURE REVIEW

Hospital pharmacy services have recently significantly transformed towards patient-centered services (Ramachandram et al., 2022) by providing value-based services to improve patient satisfaction. For example, RDS adds valuable support by handling tedious tasks, such as writing patient prescriptions, sorting, and delivering medications using a barcode system (Sinha et al., 2022). RDS increases staff productivity by augmenting many repetitive and labor-intensive tasks for pharmacists. Reduction in medication dispensing errors and patient-waiting times at the pharmacy are also favorable outcomes of automation (Tan et al., 2009). However, the increasing patient population is causing an increased workload and negatively affecting patient wait time, satisfaction, and medication error rate. Waiting time is affected due to a multitude of factors, such as manual dispensing and nonstandard dispensing processes. Nonstandard processes and human errors in medication dispensing can be fatal and may lead to patient dissatisfaction.

The timeliness of the preparation of medications contributes to the satisfaction of patients and nursing staff. Thus, automation in medicine dispensing procedures could reduce the time spent preparing the medication supplies and patients' waiting time, which is a significant setback in a hospital pharmacy (Selvaraj et al., 2022). The study conducted in an Outpatient Pharmacy of a Comprehensive Cancer Center concluded that process improvement in an outpatient pharmacy effectively reduces patient waiting time and improves patient and employee satisfaction. Patient wait time significantly decreased for fewer than three medications and prescriptions of three or more medications (22.3 minutes vs. 8.1 minutes, $P < .001$-, and 31.8 minutes vs. 16.1 minutes, $P < .002$, respectively). At the same time, patient satisfaction increased (62% vs. 69%; $P= 0.005$) after full implementation of the project (Hammoudeh et al., 2020). The study by Steven Silverstein stated that Data collected over two years on robot efficiency, changing pharmacy technician roles, and pharmacist time led to the multiple iterations of the medication distribution process and pharmacy technician roles to increase workflow efficiency from 71% to 90% cart (Silverstein et al., 2010).

Every Year, 7,000 to 9,000 people die in the USA from medication dispensing errors (Tariq et al., 2022). The cost of errors from medication dispensing has exceeded $40 billion annually (Tariq et al., 2022). The lack of patient reporting and subsequent complications from incorrect medication leads to adverse psychological and physical experiences for patients (Tariq et al., 2022) and results in patient dissatisfaction. The U.S. Food and Drug Administration (FDA) receives more than 100,000 reports per year on medication errors in the U.S. This affects over 7 million adults.[1] Harmful results of a medication error may include death, life-threatening health, hospitalization, disabilities, and congenital disabilities. The impact of medication error is significant. Therefore, research into the impact of dispensing errors, patient wait time, and subsequent patient satisfaction is necessary. The first challenge is that many surveys in large American hospitals have suggested that, on average, less than half of dispensing medication errors get reported. (Ehsani et al., 2013; Stratton et al., 2004). Another issue is a dearth of hospitals using robotic dispensing systems, which made it difficult to find patient or hospital feedback on RDS. Hence, efforts are required from both public and private hospitals to conduct more studies and analyze the already-available RDS data variables.

Implementing RDS is a significant investment and requires substantial changes to hospital infrastructure (Alahmari et al., 2022). The effect of pharmacy automation on inventory control, billing, workload, potential medication errors, and medication dispensing time has been researched in the extant studies; however, the impact of medication dispensing rate on patient

[1] https://www.fda.gov/drugs/information-consumers-and-patients-drugs/working-reduce-medication-errors

waiting time is inadequately studied (Chua et al., 2009). Extant evaluation studies on RDS lack sufficient pre- and post-implementation data and significant results/metrics (Boyd et al., 2018). Consequently, an in-depth evaluation of RDS for pharmacy automation is necessary to inform hospital executives and make an educated decision to implement RDS. At the very least, the results should offer prospective hospitals the opportunity to research the viability of RDS.

THEORETICAL MODEL AND METHODOLOGY

The following hypothesis is tested to evaluate the RDS (Please see Figure 1)
H1: RDS implementation reduces patient wait time.
H2: RDS implementation reduces the prescription error rate.
H3: Wait time negatively impacts patient satisfaction.
H4: Dispensing error rate negatively impacts patient satisfaction.

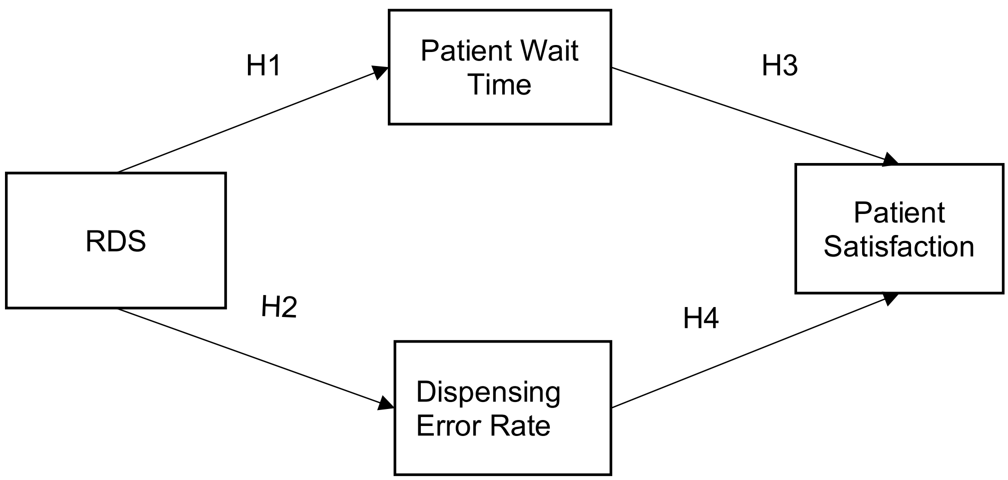


Figure 1: Study Hypothesis - Impact of Remote Dispensing System

This study tests the above model in a hospital that adopted RDS. The hospital had a traditional pharmacy with a manual medication dispensing system. RDS was launched in April 2019. For RDS installation, the entire manual pharmacy process flow was revised. The hospital management information system (HMIS) was integrated with the RDS technical system to accommodate a revised dispensing process. The current study acquires pre- and post-RDS implementation data on medication dispensing error rate, patient wait time, and satisfaction. Dispensing error rate and Patient wait time data are generated by the system. Patients were requested to provide feedback regarding their satisfaction. Subsequently, data collected over 24 months are used for evaluating the impacts of the RDS on dispensing error rate, patient wait time, and, eventually, satisfaction. The monthly data collection was done from April 2018 through April 2020. The data collection period is divided into 12 months before (May 2018- April 2019) and 12 months (May 2019 - April 2020) after the RDS system's implementation. Please refer to Figure 2 and Figure 3 for the number of prescriptions and respondents pre-post RDS.

Monthly dispensing error data is collected in three categories: minor, moderate, and major. The patient wait time is measured starting from the patient receiving the token when entering the outpatient pharmacy and to the point where the token is returned after the medicine is dispensed to the patients. Patient wait time is recorded in real-time, and the hospital's IT staff extracts the patient wait times from HMIS. The hospital used a structured feedback form for the patient satisfaction data collection.

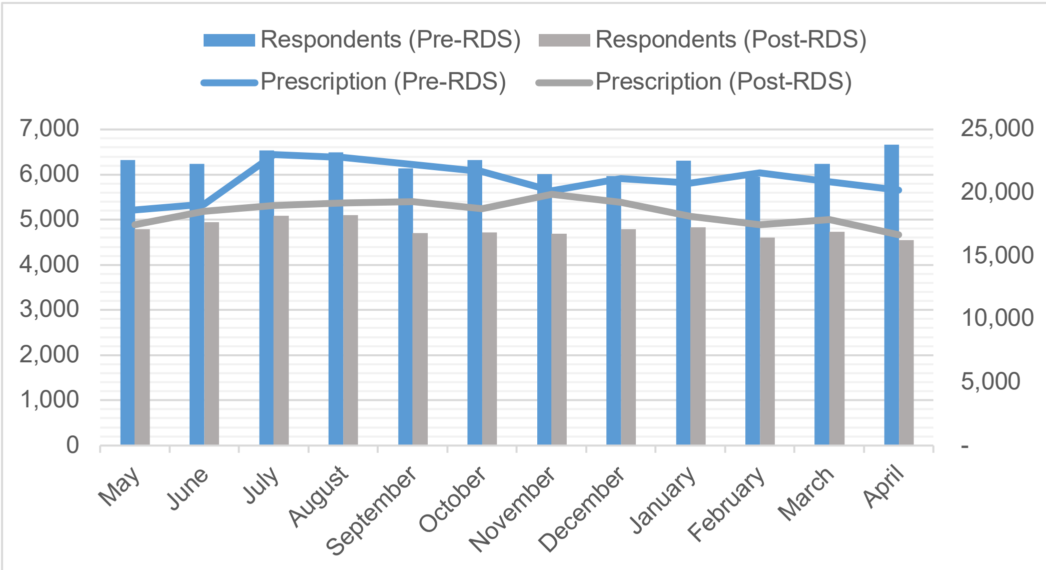


Figure 2: PRE & POST RDS | Prescription and Respondents per Month

The dependent variable is RDS implementation, and the independent variables are the medication dispensing error rate, including all categories (major, moderate, and minor), patient wait time, and patient satisfaction. The descriptive summary analysis, t-test, and SEM are applied for data analysis and result interpretation. The advantage of pre-test and post-test experimental design with real treatment is that it reveals the actual effect of RDS implementation. Also, the SEM method was selected due to the natural dependency of variables and previous findings about the mediator effect of waiting time on satisfaction (Djelassi et al., 2018). Subsequently, the significance and effect of RDS on the putative hypotheses are defined by the t-test.

RESULTS

Please refer to Table 1, Figure 2 - Figure 5 for a descriptive summary and analysis of data collection. Table 1 summarizes the key variables (Error/Rate, Wait time, Satisfaction). The total number of prescriptions reduced with average monthly prescriptions for medications before the implementations was 21,005, and after the implementation, the prescriptions count was 18,451 (Refer to Figure 3). Similarly, the number of respondents (Refer to Table 1) and the average monthly number of respondents reduced (Refer to Figure 3). 2-tailed t-test for group difference for each independent variable resulted in a favorable significance value ($p>0.0001$). The total number of errors reduced from twenty-six (26) Total Errors to five (5) Total Errors, with a medication error rate from 0.103% to 0.022% in every 1000 patients (Refer to Table 1). Also, there was no incident of major medication error post-RDS implementation for 12 months. The patient wait time was reduced from 35.62 minutes to 17.29 minutes per

2092175

prescription, a 50% reduction in wait time. Patient satisfaction increased from 68.45 % to 84.51 %, a 23% rate of satisfaction improvement. Medication errors, reducing wait time, and improving patient satisfaction.

Table 1: Summary - Data Collection

	Pre-RDS (May 2018 - April 2019)	**Post-RDS (May 2019 - April 2020)**	**Average**	**Std. Deviation**
Total Errors	26	5	15.50	14.85
Major Error	2	0	1.00	1.41
Moderate Error	6	2	4.00	2.83
Minor Error	18	3	10.50	10.61
Patient Wait Time (min.)	35.42	17.29	26.35	9.38
Dispensing Error Rate	0.0103	0.0023	1.29	1.267
Patient Satisfaction (%)	68.45	84.51	76.48	8.66
Total Prescription	252,005	221,412	236,709	21,633
Respondents (N)	75,187	57,564	66,376	12,461.
Respondents-Prescription Ratio	0.30	0.26	0.28	0.03

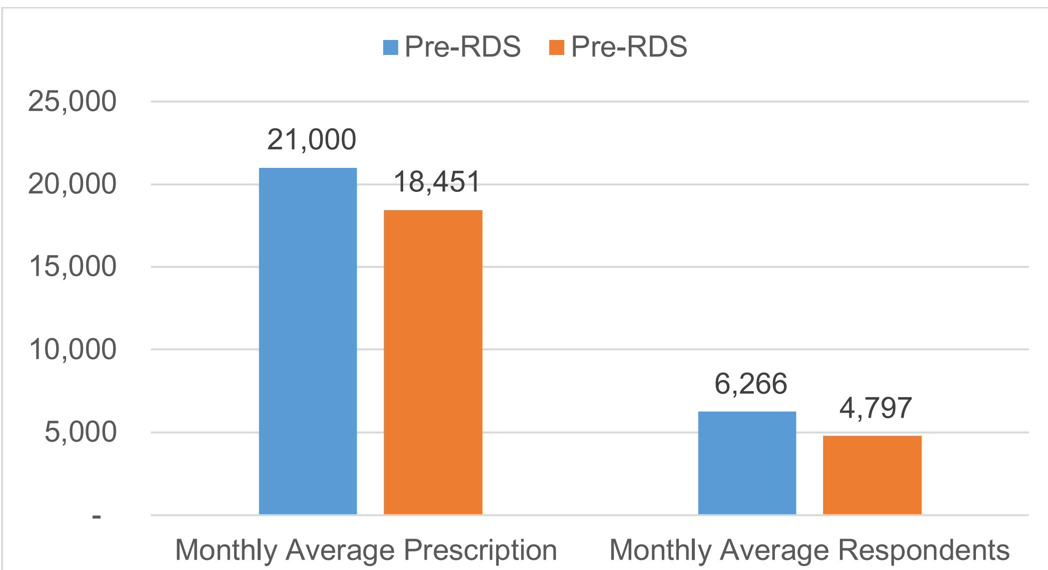


Figure 3: Pre & Post RDS | Prescription and Respondents (Monthly Average)

2092175

Figure 4: Box and Whisker - Satisfaction & Wait Time

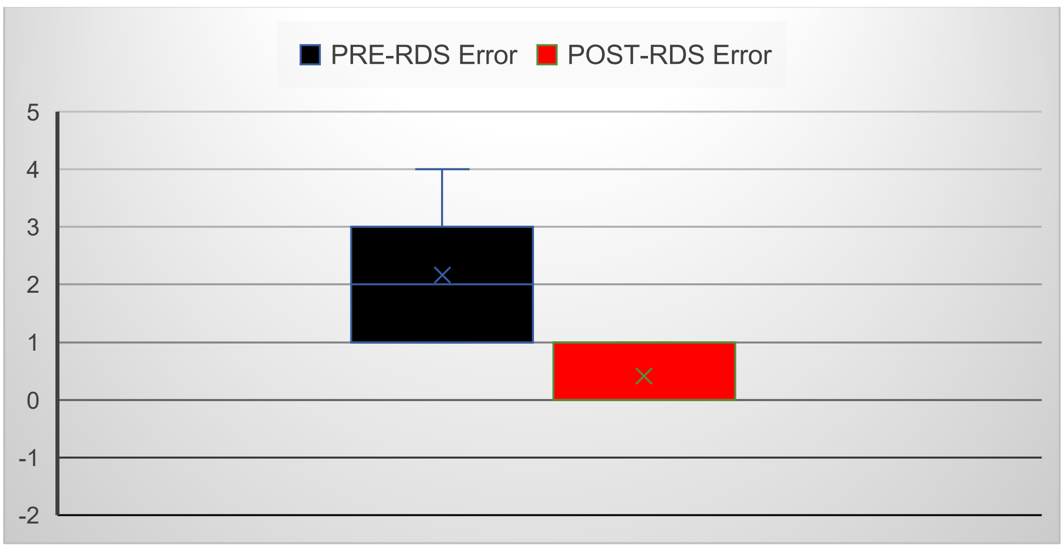


Figure 5: Box and Whisker - Dispensing Error

Box and Whisker Plot (Figure 4 & Figure 5) show the distribution and median values of the independent variables. It is clear from the visuals that the median values of variables' pre- and post-RDS median values are significantly different. Spread does not overlap and suggests the impact of RDS.

Table 2: Independent Variables- Standardized Coefficient and Standard Error

Relationship	**Std. Coefficient**	**Std. Error**
RDS implementation -> Wait time	-0.98***	0.004
RDS Implementation -> Dispensing Error Rate	-0.70***	0.074
Wait Time -> Satisfaction	-0.92***	0.064
Dispense Error Rate -> Satisfaction	-0.001	0.080
* p < .10; ** p<.05; ***p<0.01		

2092175

SEM results are summarized and displayed in Table 2 and Figure 6. RDS implementation has an inverse and significant (P<0.0001) relationship with the patient wait time (-0.98) and medication dispensing error rate (-0.70). In contrast, the patient wait time has an inverse (-0.92) and significant (P<0.0001) relationship with patient satisfaction. Furthermore, patient wait time has an inverse (-0.001) but not significant (P<0.080) relationship with patient satisfaction.

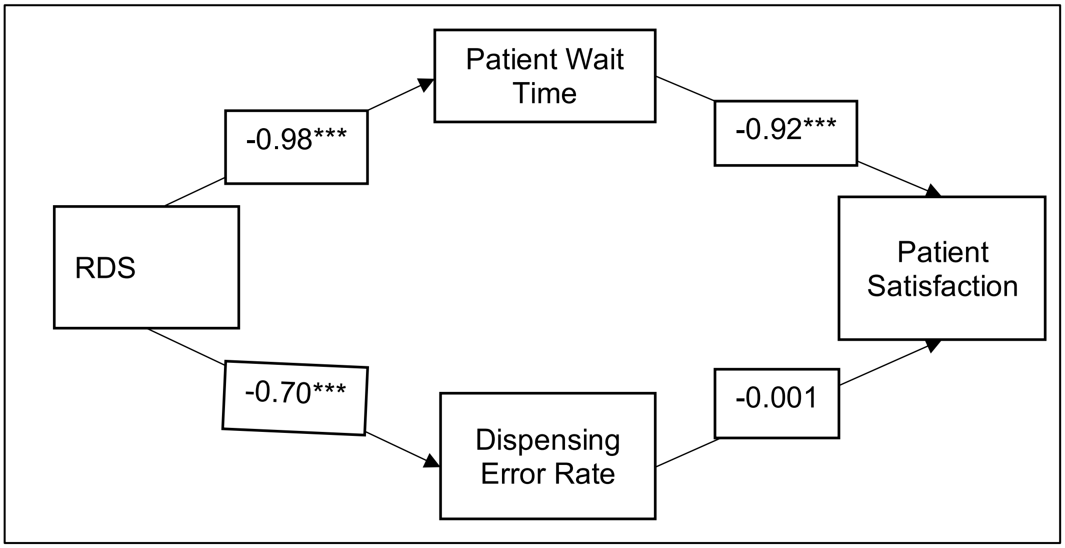


Figure 6: Impact of Robotics Dispensing System

DISCUSSION

The results from both the descriptive and SEM models support the study hypothesis. The drop in prescription numbers is attributed to an overall change in implementation and/or the COVID-19 pandemic. Similarly, the percentage of patient feedback reduced and could be attributed to increased satisfaction. The effect of RDS on patient wait time and dispensing error is significant. Similarly, reduced patient wait time has significant effects on patient satisfaction. However, the effect of the dispensing error rate is not significant. It can be argued that patient wait time affects satisfaction more than dispensing rate; consequently, patient satisfaction is biased towards patient wait time.

The findings indicate that the deployment of RDS has a beneficial effect on patient satisfaction. RDS implementation has increased overall patient satisfaction with pharmacy services. External factors such as social distancing, fewer appointments, and fewer in-person hospital visits during COVID-19 did not affect the satisfaction rate. This could be important for executives and decision-makers as factors such as COVID-19 directly impact first responders, patient service, and the hospital's overall economic sustainability.

Overall, the dispensing error rate decreased. However, major errors were reduced to zero within the first year after RDS adoption. Reduction in error improves outcomes quality and will improve overall quality metrics of the hospital (Donabedian, 1968) and, in turn, can improve the incentive under value-based care (VBC) program (*CMS' Value-Based Programs | CMS*, 2022). The process change has eliminated several manual tasks that the technicians and pharmacists used to perform. The new RDS system has provided pharmacists more time with each patient for counseling instead of looking up medications and performing administrative duties at the pharmacy.

However, there are unintended consequences related to pharmacy automation and robotics systems (Boyd et al.,2018), such as increased maintenance and the need for significant capital infrastructure. The complexity of various pharmacy procedures necessitates specific technology implementations for every institution. Pharmacy leaders need to be aware that implementations are not universal, and that each leader needs to be aware of the quirks of their institution, even while it is crucial to learn about experiences at various institutions.

Due to emerging technologies, financial incentives, and financial pressures, health care is continually evolving—thorough assessment by pharmacy automation and robotics solutions leaders before the introduction is frequently absent. Therefore, pharmacies are under pressure to perform rapid-cycle evaluations and implement solutions to satisfy rising expectations to increase effectiveness, improve patient satisfaction, reduce wait time, and reduce the dispensing error rate. Consequently, RDS implementation shall be pursued after careful consideration backed by research studies. This study contributes to the growing research on the drug industry's distribution and optimization. Our findings provide valuable insights into the technological impact on the distributors and their end clients. The findings highlight the areas in which pharmacies could take action to improve performance and patient satisfaction.

CONCLUSIONS

Due to the need for more research on the impact of RDS, pharmacy administrators frequently choose ways to maximize pharmaceutical use within their facilities based on marketing claims and anecdotal information from colleagues rather than seeking proof that these claims are supported. The results and conclusions from the current study shall provide valuable insights to hospitals to evaluate the feasibility of the RDS system. The factors considered in the current study, patient wait time, dispensing error rate, and satisfaction, are crucial for any healthcare service to retain happy customer satisfaction. Other healthcare facilities can consider the study's findings throughout the RDS planning stage. When the RDS technology element is upgraded, more research and analysis may be done to determine how it lowers medication errors, improves patient wait times, and enhances satisfaction.

ACKNOWLEDGMENT

We humbly acknowledge the support of the private hospital in Dubai, UAE, for the approval of using the patients and hospital data in this study and the enormous efforts given by their team to support this research work.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFERENCES

Alahmari, A. R., & Alrabghi, Khawlah K. (2022). An Overview of the Current State and Perspectives of Pharmacy Robot and Medication Dispensing Technology. Retrieved from https://www.cureus.com/articles/103287-an-overview-of-the-current-state-and-perspectives-of-pharmacy-robot-and-medication-dispensing-technology#!/

Boyd, A. M. (2018). Critical Evaluation of Pharmacy Automation and Robotic Systems: A Call to Action, 54(1). Retrieved from https://journals.sagepub.com/doi/10.1177/0018578718786942

CMS. (2019). Working to Reduce Medication Errors. U.S. Food and Drug Administration. https://www.fda.gov/drugs/information-consumers-and-patients-drugs/working-reduce-medication-errors

CMS' Value-Based Programs | CMS. (2022). Cms.gov. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs

Claire, C., & Matthieu, R. (2010). Automated drug dispensing system reduces medication errors in an intensive care setting, P 2275–2281.

Djelassi, S., Diallo, M. F., & Zielke, S. (2018). How self-service technology experience evaluation affects waiting time and customer satisfaction? A moderated mediation models. *Decision Support Systems, 111*, 38-47.

Donabedian A. (1968). The evaluation of medical care programs. *Bulletin of the New York Academy of Medicine*, *44*(2), 117–124.

Ehsani, S. R., Cheraghi, M. A., Nejati, A., Salari, A., Esmaeilpoor, A. H., & Nejad, E. M. (2013). Medication errors of nurses in the emergency department. *Journal of medical ethics and history of medicine*, *6*, 11.

Franklin, B. D., & Grady, K. O. (2008). An evaluation of two automated dispensing machines in a UK hospital pharmacy. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1211/ijpp.16.1.0009

Hammoudeh, S., & Amireh, A. (2020). The Impact of Lean Management Implementation on Waiting Time and Satisfaction of Patients and Staff at an Outpatient Pharmacy of a Comprehensive Cancer Center in Jordan, 58(6). Retrieved from https://journals.sagepub.com/doi/abs/10.1177/0018578720954147

Maria, B. A., & Luiz, B. J. (2022). Automation of a tertiary hospital pharmacy drug dispensing system in a lower-middle-income country: A case study and preliminary results. Retrieved from https://doi.org/10.1016/j.rcsop.2022.100151

Momattin, H., & Arafa, S. (2021). Robotic Pharmacy Implementation and Outcomes in Saudi Arabia: A 21-Month Usability Study, 8(July- Sept). Retrieved from https://humanfactors.jmir.org/2021/3/e28381

Rachel, O. C., Sang, W. Y., & Soongeol, K. (2021). Analysis and optimization of the replenishment process for the robotic dispensing system in a central fill pharmacy. Retrieved from https://doi.org/10.1016/j.cie.2021.107116

Rodriguez Gonanlez, C. G., Alonso, A. H., & Vilaplana, V. E. (2019). Robotic dispensing improves patient safety, inventory management, and staff satisfaction in an outpatient hospital pharmacy, 25(1), 28–35.

Ramachandram, D. S., & Kow, C. S. (2022). The Need for Automated Drug Dispensing Systems (ADDS) at In-Patient Pharmacy Departments in Malaysia: A Brief Overview, 58(2). Retrieved from https://journals.sagepub.com/doi/10.1177/00185787221122653

Sarah, B., Ana, H., & Nicolas, R. (2020). Automation of in-Hospital Pharmacy Dispensing: A Systematic Review, 28(2).

Silverstein, S. (2010). Pharmacy Automation and Workflow Implications: A Case Study, 26(2). Retrieved from https://journals.sagepub.com/doi/abs/10.1177/875512251002600203

Stratton, K. M., & Blegen, M. A. (2004). Reporting of medication errors by pediatric nurses.

Tariq R A, & Sinha, V. R. (2018, August). Medication Dispensing Errors and Prevention. Retrieved from https://europepmc.org/article/NBK/nbk519065

Tan, W. S., & Chua, S. L. (2009). Impact of Pharmacy Automation on Patient Waiting Time: An Application of Computer Simulation, 1. Retrieved from https://annals.edu.sg/pdf/38VolNo6Jun2009/V38N6p501.pdf

Tomoki, T., Norio, M., & Naoki, S. (2022). Evaluating the Safety and Efficiency of Robotic Dispensing Systems, 5–9. Retrieved from https://doi.org/10.1186/s40780-022-00255-w

DECISION SCIENCES INSTITUTE

Incorporating Problem-Based Learning (PBL) and Team-Based Learning (TBL) to Integrate Professional Skills into the Intermediate Accounting II(III) Course at Two Universities

Judith A. Sage
Sage & Sage
Email: lloydsage@aol.com

Lloyd G. Sage
Sage & Sage
Email: lloydsage@aol.com

ABSTRACT

To incorporate professional skills (e.g., communication, teamwork) into Intermediate Accounting II(III), four continuum PBL approaches (e.g., problem-centered learning) and a TBL Project were successfully employed at two universities in different regions of the country. Pre-Post Tests indicated that the TBL Project was helpful in learning financial statement analysis topics. The PBL and TBL projects/cases were located/created using the Backward Design technique by starting with the AICPA, Pathways Commission, CGMA, and IFAC professional skill sets. A student survey evaluated selected 50+ professional skills. To help students in learning, one or more of these PBL approaches can be used.

KEYWORDS: Intermediate Accounting; Professional Skills; PBL Continuum; TBL; Backward Design

INTRODUCTION

For more than three decades, there has been a call for change in the accounting curriculum and teaching approaches by several accounting committees and organizations. Recently, the accounting profession suggested that students should obtain both generic/soft and technical skills. To attain these desired skills, the International Federation of Accountants (IFAC) identified professional skills that are essential for individuals entering the accounting profession in its International Education Standard 3 (IES3), *Professional Skills and General Education* (2019). Also, the *AICPA Pre-Certification Core Competency Framework* (2019) [AICPA 2019 Framework] proposed a set of necessary skills-based competencies (both technical and generic/soft) for accounting students (public/industry/government/not-for-profit) who are beginning their professional careers. In addition, the Association of International Certified Professional Accountants in their *CGMA Competency Framework: 2019 Update* (2019) [CGMA Framework] have classified skills that are deemed necessary for managerial accountants. Further, the Pathways Commission (2015) advocated approaches or learning experiences that should inspire students to think, perform, and make decisions that are comparable to the decisions of accounting professionals. Several teaching approaches (e.g., problem-based learning) are presented in this paper that can be applied in teaching both technical and generic/soft skills in an Intermediate Accounting II(III) course.

Earlier, the Accounting Education Change Commission [AECC] (1990) stated that accounting students should actively participate in the learning process and not be just passive recipients of information. For example, the preparation of a TBL Project. This research paper investigated the effectiveness of selected PBL activities and the TBL Project teaching method that engage students enrolled in their final intermediate accounting course [Intermediate Accounting II (III)], which was taught at two universities (midwest urban and southwest). The PBL and TBL activities permitted the students to experience several of the competencies/skill sets recommended by the professional organizations (e.g., IFAC, CGMA, and AICPA).

THEORIES

Learning Theories

Taylor and Hamdy (2013) denoted that "learning" involves the acquisition of domains (i.e., knowledge, skills, and attitudes). Ideally, any learning theory should include each of these domains. Presently, adult learning theories can be divided or clustered into groups or categories (e.g., instrumental theories, humanistic theories). Further, it has been implied that learning originates with the learner's existing knowledge [e.g., Vygotsky (1997)].

Instrumental Learning Theories

Instrumental learning theories emphasize individual experiences in learning. This group includes cognitive, behavioral, and experiential theories. Cognitive theory relates well with the concept that learning begins with the learner's existing knowledge. Taylor and Hamdy (2013), however, have specified that the new knowledge must be sufficiently similar to previous knowledge to allow its relevance to be recognized.

The experiential learning theory [e.g., Bruner's (1966) and Davidson's (1990) discovery learning] is encompassed in the instrumental learning cluster. Under this theory, the students' role requires active participation (engagement) in experiences that construct their knowledge base. The experiential theory appears to be relevant to accounting education since it places emphasis on the development of competencies (e.g., intermediate accounting topic(s), problem solving) and includes skills of practicing accountants (e.g., decision making, communication) in specific situations.

Humanistic Theories

Humanistic theories support individual development. This group of theories is learner centered (i.e., self-directed learning). Self-directed learning is focused on adults who are planning, conducting, and evaluating their own learning. However, as related to students, Norman (1999) and Hoban et. al. (2005) implied that it really should be "directed self-learning" instead of "self-directed learning." Directed self-learning can motivate students to undertake more responsibility in their own learning, which could prepare students for life-long learning.

Portions of the Instrumental Learning Theories and the Humanistic Theories are encompassed in the Problem-Based Learning (PBL) teaching approach. PBL is student/learner motivated and entails directed self-learning (i.e., Humanistic Theories). PBL combines the Cognitive Theory by having the students use their existing knowledge when they identify what still needs to be acquired or learned and the Experiential Theory when the instructor coordinates the students' activities to create the desired learning experience (i.e., Instrumental Learning Theories).

Problem-Based Learning (PBL)

Problem-Based Learning encompasses the principles that learning develops from cognitive and social interactions in problem centered situations (e.g., Evensen and Hmelo 2000; Savery and

Duffy 2001). PBL is a learning technique that requires students to be actively engaged in collaborative team or group projects. Under PBL students have to take considerable responsibility for their learning. Bates *et. al.* (2013) indicated that PBL results in students becoming active learners in their learning and not passive recipients of information. This is precisely what the Accounting Education Change Commission (1990) previously endorsed. Under the PBL approach the starting point for learning is a problem or query that the students endeavor to solve according to Boud (1985). The idea of the PBL concept is that students as they work on solving the problem or query will have to identify and search for the knowledge needed in order to attempt to solve the problem. Bates *et. al.* (2013) suggested that the essential factors to attain the desired learning objectives is for both the students and faculty to understand how the learning process works and their roles in this process.

Educator's and Student's Responsibilities

In order to accelerate learning using the PBL approach, educators need to (1) create the desired learning experience (e.g., project, case, module), (2) facilitate students access to the experience, (3) organize the experience and (4) provide feedback and assessment. In developing the project/case to use in the PBL approach, Wiggins and McTighe (1998) promoted using a process called the "Backward Design" (i.e., outcome-based approach). Under this method, the starting point is to identify the desired learning goal(s) of the project/case (e.g., starting with the accounting profession's learning objectives/elements). Next, the feedback and assessment activities should be established while designing a meaningful PBL project/case.

The role of learners according to the constructive learning theory is to actively participate in activities that construct their knowledge base. Taylor and Hamdy (2013) suggested that students while engaging in a PBL project should (1) expect to have to perform some searching for needed information, (2) expect to be mentally challenged, (3) construct new knowledge, and (4) hopefully have their perception, views, and beliefs supported and/or changed.

What are the benefits of utilizing the PBL approach? Allen (1992) indicated that this approach encourages the acquisition of generic or soft competences (e.g., problem solving, communication, teamwork), which are essential skills according to the Pathways Commission Learning Objectives (2015), CGMA Framework Knowledge Areas (2019), and the AICPA Framework Core Competencies (2019 and 1999 versions). PBL can be used as a strategy to encourage deep learning by the students instead of surface learning. Further, the PBL approach should permit students to activate previous learning while permitting them to incorporate or link the new knowledge with their prior learning.

PBL Continuum of Approaches

Davis and Harden (2009) have implied that PBL is not solely one approach but rather a continuum of approaches to be used by educators in teaching (e.g., problem-assisted learning, problem-centered learning, problem-based learning). An educator can use one or more of these approaches in helping students in learning. According to Barrows (1986) the PBL approach selected to be used varies with the desired learning goals or objectives (e.g., development of self-directed learning skills; increase students' motivation to learn).

Research has suggested that instructors need to establish different learning opportunities to accomplish different types of learning objectives [Anderson (1995), Driscoll (1994), Gagné and Medsker (1996), Gredler (2009), and Schunk (2020)]. Boh *et al*. (2001) indicated that lecture-based training may not be an adequate transfer technique when complexity of knowledge is high. Bonner (1999, p. 11) suggested that “learning objectives involving complex skills require teaching methods that promote active learning on the part of the students, while learning objectives involving simpler skills can be achieved with more passive teaching methods.” Harden and Davis (1998) discussed the various PBL approaches based on the relationship between the scenario/problem and the learning that can be derived from studying that problem. These authors developed an eleven-step continuum between the problem and expected learning experience by the students (e.g., theoretical learning, task-based learning).

Team-Based Learning (TBL)

Parmelee *et. al.* (2012), stated that TBL involves active learning, which is learner-centered, that holds students accountable for preparation (e.g., financial statement ratio calculations) and engagement in group activities (e.g., discussion related to the team response) and expects students to use knowledge acquired to solve authentic or realistic problems, make decisions, and communicate conclusions. These skills are what the accounting profession has been seeking. That is, students entering the profession should be able to work on teams to solve problems, make business decisions, and present the conclusions clearly to the intended audience (e.g., oral presentation, written report, business letter) according to the AICPA Framework (2019), CPA Vision Project (2017), Pathways Commission (2015), IFAC’s IES3 (2019), and CGMA Framework (2019).

Like PBL, the educator should probably use the “Backward Design” [i.e., specific desired learning goal(s)] recommended by Wiggins and McTighe (1998) to locate or create the problem, project, or case to be used under the TBL approach. The Pathways Commission Learning Objectives (2015), the CGMA Framework Knowledge Areas (2019), the AICPA Framework Core Competencies (2019), and/or the IFAC’s IES3 Professional Skills (2019) can be the source(s) for the specific learning goal(s) to be used in developing the TBL Project.

The TBL technique along with several of the eleven PBL steps can be used to achieve the specific desired learning objectives for a course over various times during a semester. Next, the authors will discuss how they have implemented PBL and TBL in teaching the Intermediate Accounting II(III) course.

TEACHING USING PBL AND TBL

To incorporate several of the Pathways Commission’s Learning Objectives (2015), CGMA Framework Knowledge Areas (2019), IFAC’s IES3 Professional Skills (2019), and the AICPA Framework Core Competencies (2019 and 1999 versions) into student learning experiences, the authors have utilized several of the continuum PBL approaches and a TBL project in teaching Intermediate Accounting II(III). The selected PBL approaches and the TBL project that have been successfully employed in teaching Intermediate Accounting II(III) at two universities (e.g., large urban, regional state) in different regions of the country (i.e., midwest and southwest) are presented below.

Problem-Based Learning Approaches

In the PBL approaches the students begin with the problem, case, or project and then they start to pinpoint and search for the knowledge needed to be learned in order to attempt to solve the problem. While the students are solving the problem, they are also learning knowledge related to that problem. In Intermediate Accounting II(III) these approaches have been utilized in this course.

Problem-Assisted Learning Approach

During the first day of the Intermediate Accounting II(III) class, the students were introduced to PBL by using the Harden and Davis (1998) "problem-assisted learning approach." The educator selected a multifaceted problem or a series of brief exercises from the textbook to assist the students in identifying keywords or features in the problem situation that may be useful in answering the question(s) (e.g., types of bond issuances, stock options).

Next, the educator asked the students what type of bonds have been issued and what is the potential impact on earnings per share for the bonds issued (e.g., convertible bonds). Once the class decides on the appropriate topic to be investigated, the students were asked to discuss with neighboring classmates what they think is the answer(s) considering facts stated in the particular situation.

Finally, a volunteer was asked to give their conclusion and the reasoning for the answer. This approach was utilized each time to introduce the students to new topics in the subsequent textbook chapters.

Also, during the first-class session, the students were informed that they will be held responsible to read the chapter for the next class period along with preparing the assigned homework problems including justification for their answers.

Problem-Solving Learning Approach

During the second-class session, the educator introduced the students to the next step in the Harden and Davis (1998) PBL continuum (i.e., "problem-solving learning"). Under this approach, the students orally present their justification or reasoning along with the calculation procedure for each problem/situation to their classmates. This permits the students under friendly conditions to begin to improve their oral communication as recommended by the AICPA Framework (2019), CPA Vision Project (2017), Pathways Commission (2015), IFAC's IES3 (2019), and CGMA Framework (2019).

In the situation where a classmate does not agree with the answer, that classmate can give their answer with their reasoning or justification. This permits an investigation of which classmate has the better reasoning or calculation for the situation and at times a discussion of where the other student took the wrong step in the analysis. The class discussion also permits the educator to expand on the topic(s) if necessary. The problem-solving learning approach is used for most class periods during the semester.

Problem-Centered Discovery Approach

Because of the complexity of the topic to be taught (e.g., cash flow statement), there are times in a semester that another PBL approach needs to be employed in place of the problem-solving learning approach. Therefore, the problem-centered discovery approach is applied. In this situation, the educator could utilize an instructor-prepared checklist along with a complex textbook problem in class to assist the students to discover and organize a complex topic (e.g., cash flows from operating activities).

Under the problem centered discovery learning approach, the students by using the checklist should discover the cash flow classification for each transaction or situation in the problem.

Then, the students should use this checklist to discover (i.e., determine) the classification treatment (i.e., increase or decrease in cash flows) for the specific transaction before the transactions' information data is entered in the appropriate cash flow statement category (e.g., operating activities, financing activities). After all the transactions have been classified and entered under the appropriate category (e.g., investing activities) as an increase or decrease, the students can discover how each of the transactions affect cash flows. Finally, the students can discover after netting the various cash flow categories that the ending cash balance on the Cash Flow Statement is the same dollar amount balance as in the ledger Cash account. It is suggested to the students to utilize the checklist in doing their homework assignments and preparing for the quiz/exam on the topic(s).

Problem-Initiated Learning Approach

The problem-initiated learning approach is another type of PBL assignment that can be used in an intermediate accounting class. Under this approach, the problem or project should act as a trigger to start the students learning on the assigned topic(s) according to Harden and Davis (1998). This approach can be used as either a team or individual student project. This learning approach has been integrated into the Intermediate Accounting II(III) course by using a Topic Project.

Topic Project—The students are given once a semester an outside of class problem-initiated learning assignment (i.e., Topic Project) to be completed in one week by each student. The students were assigned an accounting topic [e.g., investments in debt securities; investments in equity securities and other financial reporting issues; corporate capital and reacquisition of shares; dividend policy and presentation and analysis] to organize before any class discussion on that topic.

This Topic Project was assigned to encourage the students to learn how to organize topics, which is one of the skill sets (i.e., ability to organize information) recommended by the AECC (1990) and AICPA Framework (1999). The Topic Project enabled each student to organize the topic(s) using a checklist, chart, graph, grid, flowchart, outline, or other approach that will help them understand the assigned topic(s). The AICPA Framework (1999) specified the need for accounting professionals to express information and concepts in a clear and concise written manner. As a result, the Topic Project was limited to 1 1/2 pages in length.

Team-Based Learning

As discussed earlier, the AICPA Framework (2019), Pathways Commission (2015), IFAC's IES3 (2019), and CGMA Framework (2019) have all indicated that graduating students should be able to work on teams to solve problems, make business decisions, and present the conclusions clearly to the intended audience (e.g., oral presentation, written report, business letter). However, the several problem-solving learning exercises and the problem-centered discovery learning approach may not be adequate to develop these desirable skills. Therefore, a TBL Project was assigned, which will take three to four weeks to complete primarily outside of the classroom.

The TBL Project was designed using the "backward approach" (e.g., starting with the accounting profession's learning objectives/professional skills/knowledge areas/core competencies). The TBL Project involved financial statement analysis for the two most recent years of annual statements and/or SEC 10-K reports for two companies within the same industry (e.g., Home Depot and Lowe's) for each team. Each team analyzed a different industry (e.g., hotels, farm equipment manufacturing or retail department stores). An example of the instructions for this TBL Project is available upon request.

RESEARCH METHODS

One of the limitations of educational research that is conducted at only one university is whether the results will apply to other university settings. Accounting instructors should be interested in teaching techniques or methods that might be successfully utilized in different university environments.

Intermediate accounting classes taught by one of the researchers at each of the two universities were used in the research experiment. One of the universities was a midwest urban state university and the other was a regional state university [95% English second language] located in the southwest.

An end of the semester survey was used to measure the students' perceptions of the Topic Project and TBL Project. Since the "backward approach" (e.g., starting with the accounting profession's learning objectives/professional skills/knowledge areas/core competencies) was used to develop these projects, the survey purpose was to determine the student's perception if the selected professional skills were accomplished while preparing these projects.

Topic Project

One of the learning objectives of the Topic Project was to encourage the students to learn how to organize topics. Also, this project was designed to enable each student to organize the topic(s) so it will help them understand the assigned topic(s). Since this Topic Project was limited to 1 1/2 pages in length, another objective was to help the students to learn to express information and concepts in a clear and concise written manner.

One method to evaluate the accomplishments of these learning goals is to obtain the students perceptions that these skills were achieved. Several questions on the end of the semester survey can be used to measure whether the students felt they had accomplished the desired skills.

TBL Project

As previously stated, the TBL Project involved financial statement analysis (FSA) for the two most recent years of annual statements and/or SEC 10-K reports for two companies within the same industry (e.g., Home Depot and Lowe's) for each team. Each team analyzed a different industry (e.g., restaurants, pharmaceutical manufacturing, or cereal manufacturing). Neither class had prior experience in preparing a TBL project.

The student teams were allowed to select the industry from pairs of companies identified by the instructor. After selection of the industry, each team was expected to obtain online each companies' annual reports and/or SEC 10-K reports for the two most recent years. The students formed their own teams consisting of 3-5 students. Other than the selection of the industry by the team, the entire TBL Project was designed to be prepared outside of class. However, several days before the TBL Project's due date one class session was designated as a research day to permit the teams to finish organizing and preparing their TBL Project report and oral presentation.

Written Reports and Oral Presentations

The students at both universities were required to prepare a team written report, which included four parts. First, the team was required to calculate liquidity ratios, solvency ratios, and profitability ratios for the two most recent years using the definitions of the ratios given in their textbook. Then, the students needed to use the financial statement ratio analysis results as the basis for answering a set of questions, which required them to decide what data or ratio should be used. For example, “Which company has the more favorable inventory turnover?” Also, for the most recent year, each team was required to answer another set of questions related to each company’s annual and SEC 10-K reports. For example, “What are each company’s basic and diluted earnings per share?” Finally, for the most recent quarter, the team was asked to answer a set of questions related to each company’s Form 10-Q Reports (interim reports) filed with the SEC. For example, “What is the net earnings for each company?”

In addition, to assure that each student has writing experience while preparing the TBL Project, each student was required to write a one-page report. This individual report required the students to give their opinions as to the firm they would select for investment purposes. They were expected to support their investment decisions based on the team data developed in the first four parts of the TBL Project requirements.

Also, each team was required to present their analysis in an oral presentation (15 to 20 minutes in length) to their classmates. This required the preparation of PowerPoint slides and/or other visual aids for the team presentation. Also, this involved the decision of which team member will present what part of the presentation as each team member was required to participate in the presentation. The students were expected to dress as if they were presenting to clients. The other classmates were expected to act as the client and were encouraged to ask questions of the presenting team.

Evaluations

As the Pathways Commission (2012) recommended, the TBL Project was designed to promote deep engagement of the students by holding them accountable to the instructor and fellow students through the use of an evaluation form during their oral presentations. Also, the IFAC in IES3 (2019) indicated that students should be able to scrutinize their own work through feedback from others. In addition, Stone and Lightbody (2012) found that students should learn to listen. Our TBL Project required the students to not only prepare a written report and make oral presentations, but also, to encourage them to learn to listen. Each student’s presentation was evaluated as well as the team as a whole.

The evaluation form indicated that the students would be graded (5-1 Likert scale with 5 being the best) based on (1) the content of their segment or topic presented; (2) the organization of their presentations; (3) the use of visual aids during their presentations; and (4) the effectiveness of the delivery of their presentations. These grading characteristics were defined on the evaluation form. For example, the characteristics for effectiveness of the delivery of the individual student’s presentation were listed as: (1) Did the presenter’s voice enhance the effectiveness of the presentation? (2) Did the presenter maintain good eye contact? and (3) Did the presenter sound natural and professional? In addition, the evaluation form defined the characteristics on which the teams were evaluated. The team score was based on (1) the content of the team project presented; (2) the organization of the entire team presentation; (3) the coordination of visual aids used by the team; and (4) the cohesiveness of the team presentation.

The instructor evaluated the teams and each individual team member. Also, the non-presenting class members were required to complete the evaluation form and sign the peer review evaluation form related to the other teams' student presentations (i.e., perform a peer evaluation). Each classmate and the instructor had equal evaluation weight in determining the student scores for the presentations. This helped the students to consider the presentations more seriously.

The students were informed to keep the evaluation forms confidential and that only the instructor would compute the presenter's scores. The evaluation form permitted the evaluators to write comments and suggestions for each presenter and for the overall team presentation. The instructor summarized the comments and attached them to the student's individual written report.

One of the purposes of the evaluation form was, as implied by the Pathways Commission (2015), to hold the presenting students accountable to their student peers and the instructor. Another purpose of the evaluation form was to help the students to learn to listen as recommended by Stone and Lightbody (2012). Also, as suggested by the Pathways Commission (2012), it was expected that the students would benefit from peer evaluation/observation and comments.

When the written reports were returned, each student at both universities received an evaluation form with each TBL Project's component score of 70 (i.e., team report, 40; individual presentation, 10; individual report, 15; and team presentation, 5) and the comments written by their classmates about the presentation (i.e., received a peer review).

<u>Testing</u>

One technique to evaluate students' learning is to use pre- and post-study measurements (e.g., quiz or exam). According to Angelo and Cross (1993), the purpose of utilizing pre- and post-assessment techniques is to determine whether the students have benefited from class discussions and assignments. The pre-test allows the instructor to establish a benchmark of what the students know on the subject matter being investigated before the study technique (project) is utilized.

In our study Exam II, which was administered about a week after discussing the homework problems on financial accounting ratio topics (but before the Project was assigned), was designated as the pre-test for the midwest university. Questions related to the financial statement analysis topics on the Final Exam, which was given after the students completed the Project, was considered to be the post-test for the midwest university. The results of Exam II and the Final Exam at the midwest university were used to measure the effect of the TBL Project. However, at the southwest university there was no pre-test administered.

The following hypothesis was used to test the benefit of the self-managed learning technique encompassed in the TBL Project:

> H_1: The distribution of exam scores at the midwest university on the financial statement analysis questions in Exam II (before the TBL Project) and in the Final Exam (after the TBL Project) are the same.

Student Perceptions

Ennis (1987) stated that attitudes strongly determine the ability to apply intellectual skills (e.g., written communication skills). Stout and Rebele (1996) indicated the need to examine student attitudes toward a classroom teaching method. Also, Ennis (1987) suggested that if students do not have positive attitudes toward a teaching method, desired learning outcomes (e.g., intellectual skills) may not occur.

In their study Stone and Shelley (1997) used questionnaires to measure student perceptions of the instructional processes. Ramsay *et al.* (2000) administered a questionnaire to determine students' preferences for a cooperative learning method. Sawyer *et al.* (2000) gave a survey after the students received their grade and assessment sheet for a case to determine whether the students felt that the case was useful in meeting specified skills. Chu and Libby (2010) utilized a post-assignment questionnaire to evaluate an active learning assignment.

As Stone and Shelley (1997) did in their research, our study used an end of the semester survey to measure student perceptions of the PBL and TBL techniques. This survey requested the students to rank (strongly agree = 5) whether selected Pathways Commission's (2015) Learning Objectives, CGMA Framework (2019) knowledge areas and AICPA Framework Core Competencies (2019 and 1999 versions) were achieved while doing their Topic Project and TBL Project (i.e., report preparation, team presentation, individual presentation, completion of the peer evaluation form, and reviewing peer evaluations). In addition, the survey was used to ascertain the opinions of the students on the usefulness of the TBL Project in understanding the financial statement analysis topics.

RESULTS

The students at the midwest university were tested twice on the financial statement analysis topics. First, an exam (Exam II) was administered after discussing the homework problems on these topics, but before the TBL Project (financial statement analysis) was assigned. Exam II was considered as the pre-test. The second exam (Final Exam) was given after the students had completed the TBL Project. The Final Exam was designated as the post-test.

The majority of the students' exam scores at the midwest university that were related to the financial statement analysis topics increased or stayed the same after the TBL Project was completed. At the midwest university, the median score increased from 83% on Exam II to 100% on the Final Exam. Also, the mean score at the midwest university increased from 81.17% on Exam II to 91.00% on the Final Exam.

The students' Exam II and Final Exam scores were matched by names at the midwest university. The Wilcoxon signed rank test was utilized to test H_1 (EII ≥ FE). Since there was a significant difference (at $p = .01$) for the urban midwest university, H_1 was rejected. The students' exam scores significantly increased as a result of the TBL Project at the midwest university. Thus, it appears that this type of learning exercise can be successfully used as a teaching method for financial statement analysis topics at the midwest university.

Student Perceptions

As previously mentioned, like Sawyer *et al.* (2000), the researchers administered an end of the semester survey to determine the students' perception of their PBL and TBL activities (e.g., Topic Project, TBL Project). The survey was used to give an indication of the students' belief that certain components of the Pathways Commission's Learning Objectives (2015), CGMA Framework (2019) knowledge areas, and the AICPA Framework (2019 and 1999 versions) skills were achieved while preparing the Topic Project and the TBL Project (i.e., analyzing, writing, presenting, and evaluating).

To measure the students' perception of the desired learning skills, the mean scores of their rankings (5 = strongly agree) for the Pathways Commission's Learning Objectives, CGMA Framework (2019) knowledge areas, and the AICPA Framework competencies for both universities were evaluated. Generally, the students stated that they "strongly agreed" or "agreed" that the selected learning objectives/knowledge areas/core competencies were accomplished by these Projects. The average mean score for the students at the midwest university was 4.14 and the median score was 4.00. At the southwest university the mean score was 4.25 and the median score was 4.13. The scores ranged from 4.43 to 3.57 for the midwest university and 4.50 to 3.56 for the southwest university.

Topic Project

Several of the selected learning objectives/knowledge areas/core competencies on the end of the semester survey could be accomplished while the students are consolidating, organizing, and writing the Topic Project. This survey was used to ascertain the students' perceptions of the specific professional skills expected to be accomplished while preparing this project.

A professional skill the researchers had hoped the students would feel they had accomplished while doing the Topic Project is "Organizes and effectively displays information so that it is meaningful to the receiving party." This was one of the main reasons for this Project with a mean score of 4.43 for the midwest university and 4.13 for the southwest university. Another skill related to this project is "Selects appropriate media for dissemination or accumulation of information," which had mean scores of 4.14 for the midwest university and 4.19 for the southwest university.

Generally, the students indicated that they "strongly agreed or agreed" that the desired learning skills were experienced while preparing the Topic Project. The only exception to this was at the midwest university (3.86 mean score) for "Expresses information and concepts with conciseness and clarity when writing and speaking." However, the students generally felt that they had accomplished the learning goals that the researchers desired related to the Topic Project. Also, there was very little difference in the students' perceptions of the desired learning goals between the midwest and southwest universities on the end of the semester survey learning objectives/elements questions. As a result, it appears that the Topic Project is a good learning exercise to be used in teaching selected Intermediate Accounting II(III) course topics.

TBL Project

As previously discussed, an end of the semester survey using a 5-point Likert scale (5 = Strongly Agree) was given to investigate whether the students believed that for certain selected accounting profession's learning objectives/professional skills/knowledge areas/core competencies were accomplished while carrying out the TBL Project requirements. Specific questions on this survey (after the TBL Project was returned to the students) were also used to determine if the TBL Project desired learning goals were achieved.

Some of the skills (i.e., generic/soft) that the students felt that they had accomplished (i.e., “strongly agreed” or “agreed”) during the semester were:

- Organizes and evaluates information and alternatives (mean score: 4.43 MW – 4.19 SW)
- Identifies what needs to be measured (mean score: 4.29 MW – 4.13 SW)
- Transfers knowledge from one situation to another (mean score: 4.29 MW – 4.00 SW)
- Commits to achievement of common goals when working on a team (mean score: 4.14 MW – 4.31 SW)
- Uses interpersonal skills to facilitate effective interaction (mean score: 4.14 MW – 4.19 SW)

Generally, the professionals believe that accounting students should be able to communicate. The TBL Project was designed to increase student communication skills, which is not normally part of the final semester of an intermediate accounting course. The survey had six question associated with communication skills as recommended by the Pathways Commission (2015). The majority of the students’ scores for both universities were 4.00 or higher with a mean score of 4.10 for the midwest university and 4.13 for the southwest university.
Another skill set that professionals desire students to be capable of doing is analytical thinking and solving problems. The survey questions had seven questions that examined analytical thinking [i.e., Pathways Commission (2015)]. Again, the majority of the students’ scores were 4.00 or higher with a mean score of 3.94 at the midwest university and 4.21 at the southwest university.

Because the TBL Project involved teams, the students had some leadership opportunities. Under this competency the students’ opinions of the TBL Project associated with leadership [i.e., Pathways Commission (2015)] were evaluated using 16 questions. Most of the students’ scores for these questions were 4.00 or higher with a mean score of 4.09 (midwest university) and 4.11 (southwest university).

The end of the semester survey also requested the students to rank (strongly agree = 5) whether preparing the TBL Project greatly assisted them in understanding the financial statement analysis topics. The students strongly agreed/agreed (4.50 at the midwest university and 4.38 at the southwest university) that the TBL Project increased their understanding of these topics. Finally, the students indicated on the survey (strongly agreed/agreed) that the TBL Project assisted them in preparing for the Final Exam (4.00—midwest university and 4.13—southwest university). It appears that the students at both universities found the TBL Project to be beneficial.

It noteworthy that there was very little difference in the students’ perceptions of the TBL technique between the midwest and southwest universities related to the desired learning objectives/elements questions. Therefore, it appears that the TBL Project (i.e., financial statement analysis) is a good alternative method of teaching these topics in different environments.

The TBL Project can be accomplished using a minimum of two to three hours of class time. In conclusion, it appears that this engaged learning exercise (TBL Project) can be successfully used as a teaching tool for financial statement analysis topics in the Intermediate Accounting II(III) course.

SUMMARY

The majority of the students at the midwest university received higher scores on the Final Exam related to the financial statement analysis topics after they completed the TBL Project than they did on Exam II, which reflected only regular class discussion and homework assignments on these topics. There was a significant difference at the midwest university at p = .01 between the mean scores on Exam II (before the Project) and Final Exam (after the Project). Remember that the pre-post technique was not available at the southwest university.

Many specified learning objectives/knowledge skill areas/core competencies were accomplished when the students prepared, presented, and evaluated their TBL Projects. On the survey the students ranked 74.51% (midwest university) and 86.54% (southwest university) of the learning objectives/knowledge skills/elements that the researchers expected to be achieved as either "strongly agree" or "agree." The average mean score on the end of the semester survey was 4.00 (midwest university) and 4.06 (southwest university). The median score was 4.00 for the midwest students and 4.13 for the southwest students. The scores ranged from 4.43 to 3.57 (midwest university) and 4.50 to 3.56 (southwest university).

The PBL and TBL techniques gave the students an opportunity to incorporate selected learning objectives/knowledge skill areas/core competencies that may not normally be accomplished in the typical final semester of the Intermediate Accounting II(III) course. Further, the students indicated that they "strongly agreed" or "agreed" (4.50—midwest university and 4.38— south-west university) that the TBL Project increased their understanding of the financial statement analysis topics.

Also, in the preparation of the Topic Project and TBL Project, the students were active participants in the learning process as suggested by Schulman (2005) and the AECC (1990). The use of actual company data in the TBL Project permitted the students to analyze and interpret real world financial information as advocated by Albrecht and Sack (2000). Also, as recommended by both the Pathways Commission (2015) and the AECC (1990), the students were required in the TBL Project to locate, obtain, and organize/analyze financial and nonfinancial information.

In addition, our PBL and TBL activities gave the students an opportunity to improve their communication skills as recommended by the CGMA Framework (2019), the AICPA Framework (2019), the IFAC in IES3 (2019), the Pathways Commission (2015and PricewaterhouseCoopers in Educating for the Public Trust (2003).

The TBL Project can be accomplished using a minimum of two to three hours of class time. In conclusion, it appears that the TBL Project can be successfully used as a teaching method for financial statement analysis topics in an Intermediate Accounting II(III) course. Further, since there are definitely several PBL approaches that can be easily integrated into Intermediate Accounting and other accounting courses, why not try using one or two of these approaches next semester?

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

Innovating toward CSR: Creating Value by Empowering Employees, Customers and Stockholders

Stephen K. Callaway
The University of Toledo
stephen.callaway@utoledo.edu

Narges Mashhadi Nejad
The University of Toledo
Narges.MashhadiNejad@rockets.utoledo.edu

ABSTRACT

This paper examines innovation through corporate social responsibility (CSR), which among other things, encourages corporations to represent more stakeholder interests than just those of stockholders. Limitations of the traditional corporation relate to internalization and hierarchy, as the traditional goal of the corporation is to maximize internalized value, perhaps even at the expense of its customers and employees. However, a problem with stakeholder capitalism may occur when value and power are too dispersed, and value-creation is not measured. To address these problems, the organization must innovate how it organizes, and engage its most critical stakeholders - its employees and customers.

KEYWORDS: Corporate Social Responsibility (CSR), Stakeholder Capitalism, Network Organizations, Balanced Scorecard, and Social Innovation

INTRODUCTION

Innovation is critical for an organization. According to Francis & Bessant (2005), organizational innovation can be viewed as the 4 Ps: Product innovation, process innovation, position innovation, and paradigm innovation. That is, innovation may target product improvement, process improvement, redefinition of the positioning of a product or organization, or most profoundly, the redefinition of the dominant paradigm of an organization (Francis & Bessant, 2005). Yet perhaps the most impactful form of paradigm innovation is not just to redefine the paradigm of a particular firm, but to innovate the very purpose of the firm in general. That is, the purpose of the firm has traditionally been to maximize shareholder value. However, a more contemporary perspective on the purpose of the firm is considered to be to maximize stakeholder wellbeing, meaning not just representing shareholder interests, but other key stakeholders as well. As such, stakeholder theory is a key element of corporate social responsibility (CSR).

This debate over the function of the firm relates to the value maximization proposition versus stakeholder theory (Jensen, 2002). Value maximization argues that managers should make all decisions in order to increase the total market value of the firm. Total value constitutes the sum of all financial claims on the firm, including equity, debt, preferred stock, and warrants. Alternatively, stakeholder theory argues that managers should make decisions while considering the interests of all stakeholders in a firm. Stakeholders include all individuals or groups who can substantially affect the welfare of the firm, not just the financial claimants, but employees, customers, communities, and governmental officials as well (Jensen, 2002). Similarly, Archie Carroll, advocating CSR, suggested that firms have four responsibilities: Economic, legal, ethical and discretionary, as opposed to economist Milton Friedman who suggested firms only have economic and legal responsibilities (Carroll 2004).

As arguments against the value maximization proposition, some of the primary limitations of the traditional corporation relate to internalization and hierarchy. The goal of the corporation is to maximize internalized value, perhaps even at the expense of its most critical stakeholders. For example, bargaining power of suppliers and customers of Porter's Five Forces model emphasizes this approach (Porter, 1980). The firm endeavors to achieve substantial market power, so that it can maximize and internalize value created. Market power, specifically monopoly power over its customers and monopsony power over its employees, enables the firm to maximize prices that it charges its customers, and minimize wages it pays its workers. Such a corporation may also possess a greater degree of hierarchy, i.e., control centralized with just a few executives. The end result is that substantial wealth and power reside in the traditional corporation, and with a few executives within the corporation.

On the other hand, the organization pursuing stakeholder capitalism, an important element of CSR, may organize in a way that disperses more of its value and power among more of its stakeholders. Stakeholder capitalism is a management approach that focuses on the interests of all of the organization's stakeholders, including customers, suppliers, employees, shareholders, and local communities. Following this approach, the organization's primary objective is to create long-term value, rather than maximize shareholder value by enhancing profits at the expense of other stakeholders (Schwab 2021). However, it may be difficult for organizations pursuing stakeholder capitalism to balance the conflicting goals of generating economic efficiency while simultaneously providing for their critical stakeholders. A problem with these organizations occurs when value and power are too dispersed, and value-created is ill-defined and/or not measured. Stakeholder capitalism could betray economic efficiency in exchange for participation for participation's sake. Another problem may be the emphasis on this 'larger' group, which includes stakeholders who have far less 'skin in the game' compared to employees, managers and investors. When controlled by even the most distant stakeholders, accountability to all creates accountability to none, leading to stagnation. Such organizations may be unable to make timely decisions and fail to create much value, instead emphasizing process over results. This criticism of stakeholder capitalism is similar to the criticism leveled by Jensen (2002), that any theory of action must tell the actors how to choose among multiple competing and inconsistent constituent interests. Perhaps a solution lies in focusing not on all stakeholders, but just those with the greatest stake in the organization, such as employees and customers as well as stockholders, and adopting a comprehensive strategic measurement system and strategic control system like the Balanced Scorecard (Kaplan and Norton, 1996). Such an organization would have to be innovative in the way it organizes, so as to increase participation among its employees and customers.

NETWORK ORGANIZATION

In today's competitive marketplace, organizations must be continually innovating, even including innovating their own organizational structure. According to Mintzberg, organizational structure refers to the framework of relationships among jobs, systems, operating processes, people, and groups that take action to achieve goals (Mintzberg 1979). The rigidness and hierarchy of traditional organizations are no longer sufficient for competing effectively in the business environment of today. Specifically, organizations face significant challenges because of increasing environmental uncertainty (Haarhaus and Liening 2020). Several factors have contributed to an environment that has a high level of uncertainty, including globalization, highly disruptive business models, and rapid technological change (Grant 2003, Burton 2013, Vecchiato 2015). New types of organizations are required to become and remain competitive in this new competitive and dynamic landscape, requiring strategic flexibility. In environments like these, customers' needs, and the technologies used to address them can change dramatically over time (Hitt, Keats et al. 1998). Organizations need to exercise strategic leadership, build dynamic core competencies, focus and develop human capital, use new manufacturing and information technologies effectively, and employ valuable strategies to take advantage of changing markets (Hitt, Keats et al. 1998). Organizational structure, specifically its level of organicity, is essential for these imperatives. That is, to fully conceptualize organization in this environment, it is important to focus on the two dimensions of organicity: The degree of centralization and formalization (Tannenbaum and Dupuree-Bruno, 1994). Modern organizations tend to be more decentralized and more informal than their prior hierarchical counterparts.

Research has long emphasized the importance of decentralization as an element of formal structural arrangements in organizational and management literature. Accordingly, Thompson (1967) suggested that subunits should be localized (decentralized), as a way of coping with unpredictable environments (Thompson 1967). Gordon and Miller (1976) argue that the burden of decision-making at higher management levels should be eased by reassigning tasks and responsibilities to lower levels to reduce administrative complexity (Gordon and Miller 1976). Greater decentralization allows flexibility, entrepreneurship, and motivation critical to the performance of the local unit. The uncertainty associated with the unit would tend to encourage greater decentralization as the information processing requirements of so many unknown contingencies become overly burdensome (Galbraith, 1973; Govindarajan, 1988). Because of the uncertainty creating so many contingencies and dynamism necessitating flexibility and speed, decisions need to be pushed down the chain of command (Callaway and Hamilton, 2006).

Autonomy in general is defined as having the ability to make significant decisions without others' consent (Brock 2003), and in this context, refers to organizations that are divisions or subunits of larger entities, e.g., corporations or national fraternal organizations (Hall 1991). Therefore, the definition of autonomy of an organization refers to the freedom for the subunits to manage their daily activities (Datta, Grant et al. 1991). Autonomy is the opposite of hierarchical control in organizations (Harris and Holden 2001, Darr 2003).

According to Scott (1992), organizational structure also constitutes the degree of organicity, that is, the type of structure in terms of being more mechanistic or more organic. An organic, or flexible structure is more common to complex and uncertain environments. That is, the focus is not only on technically efficient structure and routines, but on creative ways of dealing with needs and uncertainty (Scott, 1992). Facing substantial uncertainty, firms may decouple technical, formal requirements and informal, social requirements (Meyer and Rowan, 1977). Therefore, the formal organizational structure, which is purposefully designed to regulate behavior toward specific goals, may be supplemented or transformed by the emergence of an informal structure (Scott, 1992).

Modern organizational forms are also emphasizing external linkages more and more (Callaway & Hamilton, 2006). These increasingly important outside linkages increase the exposure of the local unit to an uncertain organizational field resulting from the diversity of technologies, processes, knowledge and cultures. Management of these diverse institutions effectively and efficiently, while incorporating new technological architecture, requires a greater informal and decentralized bureaucratization. Emerging customer, supplier, and competitor demographics, demands and capabilities are uncertain and unpredictable. As a result, the nature of such inter-organizational exchange relations can change, dissolve, and form again quickly, and are more likely to yield informal social relationships (Aldrich and Fiol, 1994; Callaway & Hamilton, 2006; Tolbert, 1985). Ultimately, the nature of these external linkages may be continually changing, with new organizations entering and exiting the network, even while the very distinction between internal and external linkages may become blurred. That is, the local unit must be able to develop its own culture and informal relationships, work within and outside of the organization, and forge new relationships in the changing organizational field (Callaway & Hamilton, 2006).

Other dramatic changes in the structural nature of many firms are important (Schilling and Steensma 2001). Large organizations have dramatically downsized, refocused, and vertically disaggregated. They increasingly obtain goods and services, pursue complex development efforts, and exploit horizontal synergies without traditional formal hierarchy. These firms are also internally disaggregating into smaller, more autonomous units that sometimes are treated more like external subcontractors (Zenger and Hesterly 1997). That is, many large and hierarchical organizations have given way to loosely interconnected components with semipermeable boundaries (Miles and Snow 1992, Ashkenas, Ulrich et al. 1995, Ghosal and Bartlett 1997, Zenger and Hesterly 1997). In the modern world, it is no longer possible to locate production within the confines of a single firm, rather it occurs at the nexus of relationships between various parties involved in the production process (Schilling and Steensma 2001). According to proponents of these new forms, bureaucratic structures are too slow to adapt to uncertain environments, which is why these models have been developed. Achrol (1996) and Zenger and Hesterly (1997) described this transformation as follows: The mighty multi-layered organizations of the 20th century have been gutted by downsizing, vertical disaggregation, outsourcing, and eliminating management layers. Now they are being replaced by leaner, more agile firms focused on a core technology and process, interconnected with a network of strategic partnerships and alliances with suppliers, distributors, and competitors (Achrol 1996).

Various terms have been used to describe this organizational phenomenon, including virtual organizations (Churbuck and Young 1992, Chesbrough and Teece 1998), network organizations (Miles and Snow 1992, Jones, Hesterly et al. 1997), and modular organizations (Sanchez 1995, Lei, Hitt et al. 1996, Sanchez and Mahoney 1996). Nonaka and Takeuchi (1997) further referenced other new organizational forms, including adhocracy, an infinitely flat organization, a spider's web network, an inverted pyramid, a Starburst satellite, and an internal market. Overall, these terms relating to network organization generally refer to the phenomenon of a loosely coupled network replacing the traditional role of a tightly integrated hierarchy (Orton and Weick 1990). Network organization may herald a socioeconomic shift as great as the industrial revolution (Achrol 1996). Overall, these new organizations: Possess a flatter structure than prior hierarchical models, constitute a more dynamic rather than static structure, empower people to establish relationships directly with customers, and highlight the value of local skills and capabilities (Nonaka and Takeuchi 1997). Overall, characteristics common to these new organizational forms include decentralization, localization, constituting less rigid hierarchical control, and allowing more autonomous and empowered units.

One particular type of network organization is the modular organization. Modularity is defined as a general set of principles for managing complex systems (Langlois 2002). The main idea behind modules is that a system is separated into discrete subsystems, the so-called modules. This results in smaller modules and fewer elements, resulting in simple interactions within each module and complex interactions between modules (Kuntz and Vera 2007). Depending on the degree of connectivity, a system can be considered modular to varying degrees. A typical form of hierarchical governance begins to disaggregate as a result of modularization within organizations. In order to minimize complexity, firms are decomposed into relatively small autonomous operations (modules) (Zenger and Hesterly 2008). Modularization leads to a framework where modules integrate strongly interdependent tasks while the interdependence between modules is weak (Kuntz and Vera 2007). Thus, the diffusion of modular organizational forms has been promoted by the widespread efforts of large companies to re-engineer, refocus and restructure their businesses (Zenger and Hesterly 2008). The concept of modularization is compatible with organizational structures resulting from re-engineering, since the subunits in charge of partial processes can be used as modules (Kuntz and Vera 2007). Modularization should not be viewed as an independent and self-contained organizational concept. Rather, it consists of three fundamental concepts (internal market mechanisms, process orientation, and the creation of smaller units of organization), which are integral to other organizational ideas (Zenger and Hesterly 2008).

Similarly, another type of network organization is the virtual (V-form) organization. Due to the increased pace of business and the shorter product life cycle, V-form organizations have enabled more rapid responses to market opportunities (Rossen 2002). As defined by Bultje and van Vijk (1998), virtual organizations are characterized as being a network of independent, geographically dispersed organizations. Each partner contributes critical core competencies to the network. Virtual organizations also provide products and services based on innovation and with a strong customer focus (Bultje and van Wijk 1998). Team members on a virtual team are usually geographically or chronologically separated from one another (Rossen 2002). This too reflects the organic structure, with relationships between units more flexible and dynamic, even beginning to blur the distinction between internal and external units. Consequently, organizations are increasingly viewed as complex webs of governance arrangements rather than as entities with definable boundaries (Zenger and Hesterly, 1997). As such, these organizational innovations have an important underlying commonalty: Much economic activity is converging toward exchange that involves either internal (within-firm) or external (between-firm) networks of smaller, somewhat more autonomous units (Zenger and Hesterly, 1997).

The nature of how to structure this economic activity will have an important effect on the strategic control systems used to ensure that the goals of the organization are achieved. Traditional strategic control is largely considered to comprise behavioral controls or output controls (Ouchi, 1979, Snell, 1992). The decentralization of these local units, that may connect with external organizations about as much as internal units, is a key factor determining the control systems. A highly independent unit can more likely be held accountable for its own performance, because its own efforts more directly impact results. Greater independence should increase accountability and error diagnostics and interpretation (Levinthal and March, 1993), making output-based controls more appropriate (Callaway & Hamilton, 2006). Internal market mechanisms, common to the modular organizational form, reflect the organization replicating market forces within the organization.

Overall, characteristics common to all of these network organizational forms described above include having a more decentralized (localized units) and more organic structure (maintaining relationships between units that are flexible and dynamic), and that are more controlled by internal market mechanisms than by traditional hierarchy.

EMPLOYEE PARTICIPATION

In today's competitive marketplace, organizations must be continually innovating, including strategic human resources management. Changing business conditions are encouraging senior managers to give more attention to their organizations' human development. Such conditions include advances in technology, globalization, volatile consumer demand, and environmental instabilities. In such an environment, a person's ability to cope and manage is crucial to the success of a business (Garavan 1991). Specifically, the demands on organizations necessitate other changes in organizational structure, related to human resources, as more and more organizations adopt decentralization and offer more employee autonomy and empowerment.

Decentralization is defined as an organizational structure that emphasizes employee autonomy and participation in decision making (Meirovich, Brender-Ilan et al. 2007). Similarly, autonomy is defined as having the ability to make significant decisions without the consent of others (Brock 2003). Empowerment is the process of decentralizing decision-making whereby managers offer more discretion and autonomy to employees (Brymer 1991). One type of empowerment, related to job design, constitutes structural empowerment. This structural empowerment is distinct from psychological empowerment, i.e., the employee's feeling of empowerment (Ahearne, Mathieu et al. 2005, Metcalf 2018). Substantive empowerment involves various factors including increasing the abilities, responsibilities, formal authority, effective capacity as well as the involvement of skilled employees in problem solving, decision making and continuous improvement (Vidal 2007a, Vidal 2007b, Jones, Latham et al. 2013). Empowerment equates to power. Power involves the ability to mobilize a group's internal resources, including member loyalty, organization, symbols of unity, and leadership skill, in order to extend the group's claim to enterprise resources, giving it the effective authority to claim resources and make decisions (Babson 1995, Jones, Latham et al. 2013).

There has been substantial interest in employee empowerment practices in management research and practice in recent years (Lawler, Mohrman et al. 2001, Maynard, Gilson et al. 2012, Yin, Wang et al. 2019). These HR studies have centered on the perspectives of organizational behavior (OB), the resource-based view of the firm (RBV), as well as transaction cost economics (TCE) (Yin, Wang et al. 2019). As such, research suggests that empowerment practices can improve organizational performance by fostering motivation and initiative in employees, so that they can respond to changing and competitive work environments (Patel and Cardon 2010), through creating valuable and inimitable human resources (Jiang, Lepak et al. 2012), or by improving a firm's cost efficiency (Yin, Wang et al. 2019). In short, employee empowerment can motivate employees psychologically, help an organization to build critical human capital resources, and improve the cost-efficiency of business operations.

Empowerment practices enhance organizational performance because they motivate employees to work harder (Hackman and Oldham 1976), create opportunities for employees to utilize their knowledge and skills (Wall, Cordery et al. 2002), and enable favorable social exchanges between employees and their employer (Sun, Aryee et al. 2007), as well as encourage employees to take initiative in their work (Cappelli and Neumark 2001, Yin, Wang et al. 2019). Accordingly, empowerment practices are often touted as best practices for improving employee productivity and organizational performance (Maynard, Gilson et al. 2012). Overall proponents argue that empowerment practices can bring psychological and strategic values to organizations, as well as cost efficiencies (Yin, Wang et al. 2019).

Employee empowerment practices indicate human resources practices that grant employees greater autonomy to perform their primary work (Robbins, Crino et al. 2002, Yin, Wang et al. 2019). Accordingly, there are three human resources practices associated with empowerment: Information sharing, autonomy through boundaries, and team accountability (Randolph 1995, Blanchard, Carlos et al. 1999, Seibert, Silver et al. 2004). Information sharing disseminates potentially sensitive information about the organization's financial performance, business unit outcomes, costs, and quality to employees. Autonomy through boundaries increases the capacity of employees to make autonomous decisions in their daily work. Team accountability utilizes self-managing teams and problem-solving groups to position teams as the focal point for decision-making autonomy and performance accountability in organizations. Central to these various forms of human resources practices involve changing the job design, including transferring power and decision-making authority from employers to employees (Robbins, Crino et al. 2002, Yin, Wang et al. 2019). Likewise, empowerment, according to Harley (1999) indicates delegation of responsibility from management to employees, non-hierarchical forms of work organization, as well as the sharing of information among different levels of organization (Harley 1999). Finally, organizational factors that promote organizational empowerment include: (1) dynamic structural framework; (2) decision-making control in the workplace (employee participation in all aspects of their professional career); and (3) information-sharing fluidity (providing employees with access to different kinds of information on the organization and allowing them to air their concerns and ideas (Kazlauskaite, Buciuniene et al. 2012).

This level of employee empowerment and participation will have an important effect on the strategic control systems used to ensure that the goals of the organization are achieved, that is, how the organization ensures employee actions further the goals of the company. Employee autonomy makes results more directly attributable to an employee's actions, crystalizing standards of performance (Snell, 1992). A more independent employee can more likely be held accountable for his/her own performance, with less dependence on others' actions or supervisor direction. Furthermore, disaggregation of a corporation into smaller firms or autonomous internal units enables access to high-powered incentives. These reward systems compensate managers on the basis of the performance of their subunits. With disaggregated structures as well as group-based performance assessment and rewards, employees resemble owner-operators of small, internal subcontracting units. Those internal initiatives implemented in isolation infuse market elements into hierarchy (Zenger and Hesterly, 1997). Similarly, shared capitalism compensation systems are emerging as a way to encourage employee participation. Greater involvement in shared capitalism systems has been found to lead to greater participation in decisions, better supervision and treatment of employees, more training, higher pay and benefits, greater job security and greater job satisfaction (Kruse et al., 2008). As such, employee participation overall demonstrate greater autonomy and greater utilization of shared capitalism compensation systems.

CUSTOMER PARTICIPATION

Customer participation is a term that illustrates the level of involvement of customers in the production and delivery of a service (Dabholkar 2015). Participation of customers can be divided into two components. First, customers can provide themselves with services, fully or partially, instead of relying on others to do so. For example, a customer can perform both production and consumption in the service industry, such as is the case with the self-service banking device - the automated teller machine (ATM). Second, customer participation involves gaining information at the site of service delivery, since the service provider saves time and effort by doing so (Dabholkar 2015). Moreover, the extent of customer participation constitutes two dimensions: the width of the participation and the depth of the participation (Jiang, Xu et al. 2019). A customer's participation width is the range of tasks and responsibilities that they perform as part of their participation. The more service departments and employees the participating customers must contact, the wider their participation will be (Jaworski and Kohli 1993). Depth of customer participation refers to the level of customer interdependence with the service provider(s) in the task role set. It reveals how much coordination between the customer and service provider is needed in the service process (Jiang, Xu et al. 2019).

Customer participation can be achieved by promoting customer integration into organizational processes. That is, the design and configuration of products delivered by many businesses often is determined largely by their customers. Customer integration implies the active integration of the customer in the development of goods and services in their downstream and upstream processes (Büttgen 2008). Customer integration is also important for services because they could not be provided without the customer's active engagement, presenting a huge opportunity for service firms (Chervonnaya 2003). The availability of communication technologies has enabled customers to become increasingly involved in these activities. Thus, customers become active players in the creation of value (Srivastava, Shervani et al. 1999, Sawhney 2002, Vargo and Lusch 2008, Xie, Bagozzi et al. 2008). Research on open innovation, crowdsourcing, and mass customization indicate the role that customers play in the success of companies (Chesbrough 2003, Piller, Moeslein et al. 2004, Wayne Gould 2012, Corvello and Iazzolino 2013). This approach ensures the supply process works efficiently by firms cooperating and interacting with customers (Wong and Boon-Itt 2008). Providing manufacturers and their customers with access to strategic information and fostering collaboration improves understanding of customer demand. With a better understanding of customer demands, manufacturers are able to predict market trends and make long-term operational decisions with greater accuracy (Fisher, Hammond et al. 1994).

Furthermore, according to Callaway & Dobrzykowski (2009), engaging customers directly in the decision-making of the firm represent a significant paradigm shift in the role of the firm. The old paradigm viewed value as something 'added to' products by the firm, with the customer actually destroying (consuming) this value, leaving the customer outside of the value creation system. This new paradigm argues that for value creation, one must focus downstream, to the activities of customers, acknowledging their participatory role (Porter 1985, Normann and Ramirez 1993, Vargo and Lusch 2008, Vargo and Akaka 2009). The key point is, consumers do not actually consume value, rather they are essential for devising new ways of creating value. Recognizing this fact, and organizing around it, offers substantial opportunities for innovation.

Value creation is not a linear process confined to discrete stages of a traditional value chain. Instead, it is a dynamic, networked process involving real-time interactions and feedback loops among various stakeholders, including customers and employees. In this networked model, value is co-created through the synergistic interactions among these stakeholders. Recognizing this value co-creation, firms are less focused solely on internal efficiency, and instead are increasing efforts to leverage external resources, particularly the customer, in order to create value. This requires building a new network of stakeholders, requiring them to focus on the long term and think holistically about the business. They must develop a value dense environment for customers and other value co-creating actors, characterized by high amounts of the information, knowledge, and resources. The imperative is to visualize building the whole business by constructing a complex web of stakeholder relationships, whereas these stakeholders bring resources and ideas to the network. In this more networked reality, successful firms focus efforts not only on the organization itself but on the overall value-creating system, in which different economic actors (e.g., partners, allies, customers, etc.) work together to co-create value (Normann and Ramirez 1993, Sarasvathy 2001, Prahalad and Ramaswamy 2004). Clearly the formation of decentralized networks and customer engagement are inextricably linked.

The traditional interaction between provider and customer often may have been antagonistic, due to the asymmetry of information as well as conflicting goals, resulting from the agency problem (Eisenhardt 1989). Importantly, stakeholder theory attempts to reduce or eliminate the agency problem, instead emphasizing the customer's collaborative role in value creation. Helping customers with their side of this equation, as well as redefining and reconfiguring this firm-customer interaction, can offer a promising avenue to be innovative. The critical issue is the fact that the provider may not know the latent needs of customers, while the customer may not know the operational aspects of the product. The imperative is to find ways to address the asymmetry of information and goal incongruence, to create opportunities for changing this customer-provider interaction, toward an alignment of information and goals among customers and the focal organization, which leads to the development of a value dense environment (Normann and Ramirez 1993, Vargo and Akaka 2009). So, rather than have an adversarial relationship between provider and customer (provider just trying to "sell" products, customer trying to avoid responsibility in usage problems, etc.); a mutual symbiotic relationship may be established between them, to create value for the entire network.

BALANCED SCORECARD

Measurement should be a crucial part of any CSR program. Kaplan and Norton (1992, 1993, 1996, and 2001) developed the Balanced Scorecard (BSC) in response to critiques of traditional management control. The BSC is a strategically-linked performance measurement system that emphasizes measurable objectives, targets, and initiatives. The BSC entails a "strategy map" based on causal relationships of four perspectives of financial and nonfinancial performance measurement – learning and growth, internal business processes, customer, and financial. The BSC contains outcome measures (lagging indicators) and performance drivers (leading indicators), linked together in a cause-and effect relationship. In addition to being a strategic measurement system, BSC is also a strategic control system.

The BSC's first leading indicator, learning and growth, facilitates continuous, long-term growth and improvement, brought about by investing in organizational human resources. This learning and growth process is the basis for the next perspective, internal processes. Internal process measures represent critical internal processes that address whether a firm is utilizing resources efficiently, reflecting organizational innovation and value creation to improve customer service, and hence, the next perspective. To measure the company's performance from the customer perspective, the BSC examines core metrics including customer satisfaction, customer retention, and market share, which should eventually lead to superior financial performance, perhaps measured by operating income, return on capital, and economic value added (Kaplan & Norton, 1996).

That is, the BSC expands the set of business unit objectives beyond summary financial measures. Executives now measure how their business units create value for current and future customers and how they must enhance internal capabilities and the investment in people, systems, and procedures necessary to improve future performance. The BSC captures the critical value-creation activities created by skilled, motivated organizational participants. While still emphasizing the short-term financial perspective, the BSC clearly also focuses on the value drivers for superior long-term financial and competitive performance (Kaplan and Norton, 1996). The measures are balanced between the outcome measures-the results of past efforts-and the measures that drive future performance, while also balancing between objective, easily quantified outcome measures and somewhat subjective, performance drivers of the outcome measures. Thus the advantage of the BSC is its mix of outcomes (lagging indicators) and performance drivers (leading indicators) that have been customized to the business unit's strategy (Kaplan and Norton, 1996).

THE RESEARCH MODEL

Our research focuses on three specific independent variables that best capture stakeholder theory, all demonstrating innovative models of organizational structure. These three independent variables constitute: Network Organization, Employee Participation, and Customer Participation. The dependent variable is the Balanced Scorecard (BSC).

In summary, regarding network organization,

Proposition 1: Organizations that demonstrate key characteristics of the network organization (specifically decentralized units, organic structure and internal market mechanisms) will be positively associated with superior performance, measured by the Balanced Scorecard.

Regarding employee empowerment, many organizations have been including their employees in decision-making for some time, leading to better and faster decisions, as employees are the ones who have the greatest local knowledge. Moreover, such approaches foster extrinsic motivation for employees, through superior incentives, as well as intrinsic motivation, by creating an emotional identification to the organization. Further such organizations may share a greater amount of their wealth with their employees, perhaps through higher wages and/or profit sharing. Indeed, Kruse et al., (2008) addressed shared capitalism compensation systems, or financial participation, finding that greater employee involvement in programs leads to greater participation in decisions, better supervision and treatment of employees, better training, higher pay and benefits, greater job security, and greater job satisfaction (Kruse et al. 2008). The ultimate form, and most formalized method of employee empowerment may be employee ownership of the firm. In particular, employee-owned companies (EOCs) include their own employees in more decision-making as well as firm profits. Some research suggests that EOCs boost employee productivity and satisfaction and make such firms more durable and resilient (Bryson and Freeman 2016, Walsh, Peck et al. 2018).

In the evolving landscape of Corporate Social Responsibility (CSR), the empowerment of both employees and customers stands out as a pivotal strategy. Employee empowerment, through mechanisms like higher pay, profit-sharing, and participative decision-making, grants workers greater autonomy and influence over their own working conditions. This not only fosters a sense of ownership and commitment but also acts as a safeguard against potential corporate exploitation. Similarly, customer empowerment, by giving them a voice in product development and feedback processes, ensures that product safety and quality remain paramount. Such empowerment reduces the unchecked authority of corporations, ensuring they remain accountable to both their workforce and their consumer base. By decentralizing power in this manner, the potential for corporations to exploit workers and customers is significantly curtailed, reinforcing the true essence of CSR.

Porter and Kramer (2011) introduced a concept that they termed Creating Shared Value (CSV), indicating that firms can create social value through certain business mechanisms while simultaneously collaborating with key stakeholders to make better products, improve operations, and increase financial benefits. Overall, engaging both customers and employees in the decisions of the focal organization brings two things to these stakeholders: A greater share of the profits from network operations and greater responsibility in finding innovative ways of creating that value. Specifically, redefining (improving) the use of a product or service is essential, as local customers, for example, can and should gauge the performance of the product / service over time, and help contribute to it, and share that information with providers. That is, co-creation of value between provider and customer addresses the goal incongruence and information asymmetry associated with the agency problem. Therefore networks, by expanding their focus of value creation to key stakeholders, essentially blurring the distinction between provider and consumer, can offer new ways of promoting innovation. Moreover, the measurement of this value-creation applies not just to the focal organization, but the customers and employees operating within this network as well, hence emphasizing the role of the BSC.

Proposition 2: Organizations that demonstrate key characteristics of employee participation (specifically employee decentralization and shared capitalism/financial participation) will be positively associated with superior performance, measured by the Balanced Scorecard.

Proposition 3: Organizations that demonstrate key characteristics of customer participation will be positively associated with superior performance, measured by the Balanced Scorecard.

CONCLUSION

Corporate social responsibility (CSR) is an extremely hot topic nowadays. However, the question is, is it a topic simply of discussion, in order to merely critique traditional business practices? Or is it intended to be taken as seriously as those traditional practices? Will we develop performance measurement metrics in the spirit of stakeholder theory, that are as meaningful as the traditional metrics that assess organizational efficiency and stockholder wealth creation? In the case of the latter, a contemporary strategic control system as well as innovative organizational structural forms are needed in order to foster positive and productive interactions with the organization's most critical stakeholders – customers and employees. The key is to focus on which stakeholder interactions create the most value for them and the focal organization, and measure that value. Both customers and employees should be empowered, and participate in the strategic and operational decisions of the organization. Central to this mission of better representing more disparate interests is developing new structures and processes that enable this innovative approach to CSR. The traditional, large and unwieldy, bureaucratic organizations that focused disproportionately on cost efficiency have now been gutted by downsizing, vertical disaggregation, outsourcing, and the elimination of management layers. These organizations are now being replaced by leaner, more flexible firms focused on a core technology, interconnected by a network of strategic partnerships with their most critical stakeholders (Achrol 1996). This trend toward smaller and more innovative organizational forms could also be central toward an organizational focus on many elements of corporate social responsibility – particularly stakeholder capitalism.

While this paper primarily focuses on the role of customers and employees in the value co-creation process, we acknowledge the importance of other stakeholders, such as suppliers, particularly in the context of outsourcing in large organizations. The innovative organizational form and corresponding strategic control system proposed in this paper may also be applicable to these stakeholders, and future studies could benefit from a more comprehensive examination of the value co-creation process that includes these stakeholders.

The innovative organizational form and corresponding strategic control system may help to solve a problem often attributed to CSR; that of value and power being too dispersed, value-creation being ill-defined and not measured, betraying economic efficiency, perhaps in exchange for participation for participation's sake. Further, emphasizing stakeholders who have less 'skin in the game' than the most senior employees and managers, the most loyal customers, and the most-invested stockholders, can lead to a situation of 'accountability to all creating accountability to none'. In this spirit, the focus and purpose of this paper is not simply to create a more equal distribution of the "spoils" of corporate operations, but rather to empower employees and customers to find innovative new ways of creating additional value. Perhaps then the objectives of CSR will be realized.

PRELIMINARY QUESTIONNAIRE ITEMS

Network Organization

Decentralized units (from Callaway 2006; and Gupta & Govindarajan 1991)

Organic structure (from Callaway & Hamilton 2006).

Internal market mechanisms (TRP & PRC are management tools which are implemented to substitute hierarchical coordination by internal market mechanisms. Ludwig Kuntz, Antonio Vera. 2007. Modular organization and hospital performance. Rated on a scale from 1 to 5.
-The level of use of transfer pricing (TRP).
-The level of use of profit centers (PRC).

Employee Participation

Employee decentralization (from Gavriel Meirovich, Yael Brender-Ilan, Alexander Meirovich, 2007, Quality of hospital service: the impact of formalization and decentralization, International Journal of Health Care Quality Assurance, 20(2-3):240-52.)
-The individual decision-maker has a wide latitude in the choice of means to accomplish goals
-Employees have substantial autonomy when performing their job
-Many important decisions are made locally rather than centrally
-The employees participate in the decision-making process

Shared Capitalism/Financial Participation (from Kruse at al., 2008; Lisi, 2011)
(objective data on employee pay? Higher/lower than industry average)
-Employees normally participate in profit sharing through bonuses and/or dividends
-Employees are given chances in budget preparation and execution
-Our company offers employee ownership (e.g., ESOPs) (Y or N) (new item)

Customer Participation (from Narasimham & Kim; Flynn et al.)

-The level of linkages with our major customers through information networks
-The level of sharing of market information from our major customers
-The level of communication with our major customers
-The frequency of period contacts with our major customers
-Our major customers share their demand forecasts with us

Balanced Scorecard

Internal business processes (secondary data – efficiency metrics)

Financial perspective (secondary data – profitability, returns metrics)

Customer Perspective (primary data)
-Our customer acquisition abilities (successes) are generally better than the industry average
-Our customer retention abilities (successes) are generally better than the industry average

Employee Perspective / Learning and Growth (primary data)
-Our abilities (successes) to recruit and hire top personnel are generally better than the industry average
-Our abilities (successes) to retain top personnel are generally better than the industry average

REFERENCES

Achrol, R. S. (1996). "Changes in the theory of interorganizational relations in marketing: Toward a network paradigm." Journal of the academy of marketing science 25: 56-71.

Ahearne, M., Mathieu, J., & Rapp, A. (2005). To empower or not to empower your sales force? An empirical examination of the influence of leadership empowerment behavior on customer satisfaction and performance. Journal of Applied psychology, 90(5), 945.

Ashkenas, R. (1995). The Boundaryless Organization: Breaking the Chains of Organizational Structure. The Jossey-Bass Management Series: ERIC.

Aldrich, H and Fiol, C. 1994. Fools rush in? The institutional context of industry creation. Academy of Mangement Review. 19: 645-670.

Babson, S. (1995). Lean work: Empowerment and exploitation in the global auto industry, Wayne State University Press.

Blanchard, K. H., Carlos, J. P., & Randolph, W. A. (1999). The 3 keys to empowerment: Release the power within people for astonishing results: Berrett-Koehler Publishers.

Brock, D. M. (2003). "Autonomy of individuals and organizations: Towards a strategy research agenda." International Journal of Business and Economics 2: 57.

Brymer, R. A. (1991). "Employee empowerment: A guest-driven leadership strategy." Cornell Hotel and Restaurant Administration Quarterly 32: 58--68.

Bryson, A. and R. Freeman (2016). "Profit sharing boosts employee productivity and satisfaction." Harvard Business Review 13.

Bultje, R. and J. van Wijk (1998). "Taxonomy of Virtual Organisations, based on definitions, characteristics and typology." VoNet Newsletter 2: 7-21.

Burton, R. (2013). "The future of organization design: An interpretative synthesis in three themes." Journal of Organization Design 2: 42-44.

Büttgen, M. (2008). Erscheinungsformen der Kundenintegration und Ansätze eines Integrationsmanagements. Aktuelle Forschungsfragen im Dienstleistungsmarketing, Springer: 105-132.

Callaway, S., & Dobrzykowski, D. 2009. Service-oriented entrepreneurship: Service-dominant logic in green design and healthcare. Service Science, 1(4): 225-240.

Callaway, S. & Hamilton, R. 2006. Exploring disruptive technology: The structure and control of internal corporate ventures. International Journal of Organizational Analysis. 14(2): 87-106.

Cappelli, P. and D. Neumark (2001). "Do "high-performance" work practices improve establishment-level outcomes?" Ilr Review 54: 737-775.

Carroll, A. B. (2004). "Managing ethically with global stakeholders: A present and future challenge." Academy of Management Perspectives 18(2): 114-120.

Chervonnaya, O. (2003). "Customer role and skill trajectories in services." International Journal of Service Industry Management.

Chesbrough, H. W. and D. J. Teece (1998). "When is virtual virtuous? Organizing for innovation." The strategic management of intellectual capital 27.

Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology, Harvard Business Press.

Churbuck, D. and S. Young (1992). "The virtual workplace." Forbes.

Corvello, V. and G. Iazzolino (2013). "Factors affecting the practices of external problem solvers in broadcast search." Journal of technology management & innovation 8(2): 166-177.

Dabholkar, P. A. (2015). How to improve perceived service quality by increasing customer participation. Proceedings of the 1990 academy of marketing science (AMS) annual conference, Springer.

Darr, A. (2003). "Control and autonomy among knowledge workers in sales: An employee perspective." Employee Relations.

Datta, D. K., Grant, J. H., & Rapagopolan, N. (1991). Management incompatibility and postacquisition autonomy: Effects on acquisition performance. In: P. Shrivastava (Ed.), Advances in Strategic Management (Vol. 7, pp. 157-182).

Eisenhardt, K. M. (1989). "Agency theory: An assessment and review." Academy of management review 14: 57-74.

Fisher, M. L., Hammond, J. H., Obermeyer, W. R., & Raman, A. (1994). Making supply meet demand in an uncertain world. Harvard Business Review, 72, 83-83.

Francis, D., & Bessant, J. (2005). Targeting innovation and implications for capability development. Technovation, 25, 171-183.

Galbraith, J. 1973. Designing Complex Organizations. Addison-Wesley, Reading, MA.

Garavan, T. N. (1991). "Strategic human resource development." Journal of European industrial training.

Ghosal, S. and C. A. Bartlett (1997). The individualized corporation, New York: Harper Business Books.

Gordon, L. A. and D. Miller (1976). A contingency framework for the design of accounting information systems. Readings in accounting for management control, Springer: 569-585.

Govindarajan, V. 1988. A contingency approach to strategy implementation at the business-unit level: Integrating administrative mechanisms with strategy. Academy of Management Journal. 31: 828-853.

Grant, R. M. (2003). "Strategic planning in a turbulent environment: Evidence from the oil majors." Strategic management journal 24: 491-517.

Haarhaus, T. and A. Liening (2020). "Building dynamic capabilities to cope with environmental uncertainty: The role of strategic foresight." Technological Forecasting and Social Change 155: 120033.

Hackman, J. R. and G. R. Oldham (1976). "Motivation through the design of work: Test of a theory." Organizational behavior and human performance 16: 250-279.

Hall, R. H. R. H. (1991). Organizations : structures, processes, and outcomes / Richard H. Hall. Englewood Cliffs, N.J., Prentice Hall.

Harley, B. (1999). "The myth of empowerment: work organisation, hierarchy and employee autonomy in contemporary Australian workplaces." Work, Employment and Society 13: 041-066.

Harris, H. and L. Holden (2001). "Between autonomy and control: Expatriate managers and strategic IHRM in SMEs." Thunderbird International Business Review 43: 77-100.

Hitt, M. A., Keats, B. W., & DeMarie, S. M. (1998). Navigating in the new competitive landscape: Building strategic flexibility and competitive advantage in the 21st century. Academy of Management Perspectives, 12(4), 22-42.

Jaworski, B. J. and A. K. Kohli (1993). "Market orientation: antecedents and consequences." Journal of marketing 57(3): 53-70.

Jensen, M. C. 2002. Value maximization, stakeholder theory, and the corporate objective function. Business Ethics Quarterly. 12 (2): 235-256. DOI: 10.2307/3857812

Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. Academy of management Journal, 55(6), 1264-1294.

Jiang, Y., Xu, L., Cui, N., Zhang, H., & Yang, Z. (2019). How does customer participation in service influence customer satisfaction? The mediating effects of role stressors. International Journal of Bank Marketing.

Jones, C., Hesterly, W. S., & Borgatti, S. P. (1997). A general theory of network governance: Exchange conditions and social mechanisms. Academy of management review, 22(4), 911-945.

Jones, R., Latham, J., & Betta, M. (2013). Creating the illusion of employee empowerment: lean production in the international automobile industry. The International Journal of Human Resource Management, 24(8), 1629-1645.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard-measures that drive performance. Harvard Business Review, 70(1), 71-79.

Kaplan, R. S., & Norton, D. P. (1993). Putting the balanced scorecard to work. The Performance Measurement, Management and Appraisal Sourcebook, 66-79.

Kaplan, R. S., & Norton, D. P. (1996). Linking the balanced scorecard to strategy. California Management Review, 39(1).

Kaplan, R. S., & Norton, D. P. (2001). Transforming the balanced scorecard from performance measurement to strategic management: Part 1. Accounting Horizons, 15(1), 87-104.

Kazlauskaite, R., Buciuniene, I., & Turauskas, L. (2012). Organisational and psychological empowerment in the HRM-performance linkage. Employee Relations.

Kruse, D., Freeman, R., & Blasi, J. (2008). Do workers gain by sharing? Employee outcomes under employee ownership, profit sharing, and broad-based stock options (No. w14233). National Bureau of Economic Research.

Kuntz, L. and A. Vera (2007). "Modular organization and hospital performance." Health Services Management Research 20: 48-58.

Langlois, R. N. (2002). "Modularity in technology and organization." Journal of economic behavior & organization 49: 19-37.

Lawler, E., Mohrman, S. A., & Benson, G. S. (2001). Organizing for high performance : employee involvement, TQM, reengineering, and knowledge management in the fortune 1000 : CEO report/ Edward E. Lawler III, Susan Albers Mohrman, George Benson.

Lei, D., Hitt, M. A., & Goldhar, J. D. (1996). Advanced manufacturing technology: organizational design and strategic flexibility. Organization Studies, 17(3), 501-523.

Levinthal, D. & March, J. 1993. The myopia of learning. Strategic Management Journal. 14: 95-112.

Maynard, M. T., Gilson, L. L., & Mathieu, J. E. (2012). Empowerment-fad or fab? A multilevel review of the past two decades of research. Journal of management, 38(4), 1231-1281.

Meirovich, G., Brender-Ilan, Y., & Meirovich, A. (2007). Quality of hospital service: the impact of formalization and decentralization. International Journal of Health Care Quality Assurance.

Metcalf, A. Y. a. H., Marco and Fry, Timothy D and Stoller, James K (2018). "The impact of quality practices and employee empowerment in the performance of hospital units." International Journal of Production Research 56: 5997--6014.

Meyer, J. and Rowan, B. 1977. Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology. 83: 340-363.

Miles, R. E. and C. C. Snow (1992). "Causes of failure in network organizations." California management review 34: 53-72.

Mintzberg, H. (1979). "The structuring of organizations. Engle-wood Cliffs." J: Prentice-Hall.

Nonaka, I. and H. Takeuchi (1997). "A new organizational structure." Knowledge in Organisations: 99-133.

Normann, R. and R. Ramirez (1993). "From value chain to value constellation: Designing interactive strategy." Harvard business review 71: 65-77.

Orton, J. D. and K. E. Weick (1990). "Loosely coupled systems: A reconceptualization." Academy of management review 15: 203-223.

Ouchi, W. 1979. A conceptual framework for the design of organizational control mechanisms. Management Science. 25:833-848.

Patel, P. C. and M. S. Cardon (2010). "Adopting HRM practices and their effectiveness in small firms facing product-market competition." Human Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management 49: 265-290.

Piller, F. T., Moeslein, K., & Stotko, C. M. (2004). Does mass customization pay? An economic approach to evaluate customer integration. Production planning & control, 15(4), 435-444.

Porter, M. E. (1985). Creating and sustaining superior performance. Competitive advantage, 167, 167-206.

Porter, M. E. (1980). Techniques for analyzing industries and competitors. Competitive Strategy. New York: Free.

Porter, M. E., & Kramer, M.R. (2011). Creating Shared Value. Harvard Business Review 89 (1), 4-17.

Prahalad, C. K. and V. Ramaswamy (2004). "Co-creation experiences: The next practice in value creation." Journal of interactive marketing 18: 5-14.

Randolph, W. A. (1995). "Navigating the journey to empowerment." Organizational dynamics 23: 19-32.

Robbins, T. L., Crino, M. D., & Fredendall, L. D. (2002). An integrative model of the empowerment process. Human resource management review, 12(3), 419-443.

Rossen, E. (2002). Trust and virtual organizations. IFIP World Computer Congress, TC 8.

Sanchez, R. (1995). "Strategic flexibility in product competition." Strategic management journal 16: 135-159.

Sanchez, R. and J. T. Mahoney (1996). "Modularity, flexibility, and knowledge management in product and organization design." Strategic management journal 17: 63-76.

Sarasvathy, S. D. (2001). "Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency." Academy of management Review 26: 243-263.

Sawhney, M. (2002). "Don't just relate-Collaborate." MIT Sloan Management Review 43(3): 96.

Schilling, M. A. and H. K. Steensma (2001). "The use of modular organizational forms: An industry-level analysis." Academy of management journal 44: 1149-1168.

Schwab, K. (2021). Stakeholder capitalism: A global economy that works for progress, people and planet, John Wiley & Sons.

Scott, W. 1992. Organizations: Rational, natural, and open systems. 3rd ed., Prentice-Hall. Englewood Cliffs, NJ.

Seibert, S. E., Silver, S. R., & Randolph, W. A. (2004). Taking empowerment to the next level: A multiple-level model of empowerment, performance, and satisfaction. Academy of management Journal, 47(3), 332-349.

Snell, S. 1992. Control theory in strategic human resource management: The mediating effect of administrative information. Academy of Management Journal. 35: 292-328.

Srivastava, R. K., Shervani, T. A., & Fahey, L. (1999). Marketing, business processes, and shareholder value: an organizationally embedded view of marketing activities and the discipline of marketing. Journal of marketing, 63(4_suppl1), 168-179.

Sun, L.-Y., Aryee, S., & Law, K. S. (2007). High-performance human resource practices, citizenship behavior, and organizational performance: A relational perspective. Academy of management Journal, 50(3), 558-577.

Tannenbaum, S. and Dupuree-Bruno, L. 1994. The relationship between organizational and environmental factors and the use of innovative human resource practices. Group and Organization Studies. 19: 171-202.

Thompson, J. (1967). "Organizations in action, 1967." SHAFRITZ, Jay M.; OTT, J. Steven. Classics of Organization Theory 4.

Tolbert, P. 1985. Institutional environements and resource dependence: Source of administrative structure in insitutions of higher education. Administrative Science Quarterly. 30: 1-13.

Vargo, S. L. and M. A. Akaka (2009). "Service-dominant logic as a foundation for service science: clarifications." Service Science 1: 32-41.

Vargo, S. L. and R. F. Lusch (2008). "Service-dominant logic: continuing the evolution." Journal of the Academy of marketing Science 36: 1-10.

Vecchiato, R. (2015). "Strategic planning and organizational flexibility in turbulent environments." Foresight.

Vidal, M. (2007a). "Lean production, worker empowerment, and job satisfaction: A qualitative analysis and critique." Critical Sociology 33: 247--278.

Vidal, M. (2007b). "Manufacturing empowerment?'Employee involvement'in the labour process after Fordism." Socio-economic review 5: 197--232.

Wall, T. D., Cordery, J. L., & Clegg, C. W. (2002). Empowerment, performance, and operational uncertainty: A theoretical integration. Applied Psychology, 51(1), 146-169.

Walsh, P., Peck, M., & Zugasti, I. (2018). Why the US needs more worker-owned companies. Harvard Business Review, 14-20.

Wayne Gould, R. (2012). "Open innovation and stakeholder engagement." Journal of technology management & innovation 7(3): 1-11.

Wong, C. Y. and S. Boon-Itt (2008). "The influence of institutional norms and environmental uncertainty on supply chain integration in the Thai automotive industry." International Journal of Production Economics 115(2): 400-410.

Xie, C., Bagozzi, R. P., & Troye, S. V. (2008). Trying to prosume: toward a theory of consumers as co-creators of value. Journal of the Academy of marketing Science, 36(1), 109-122.

Yin, Y., Wang, Y., & Lu, Y. (2019). Antecedents and outcomes of employee empowerment practices: A theoretical extension with empirical evidence. Human Resource Management Journal, 29(4), 564-584.

Zenger and W. Hesterly (1997). "Tbe disaggregation of corporations: Selective intervention, bigb-powered incentives and modular units." Organization Science 8: 209-222.

Zenger, T. R. and W. S. Hesterly (2008). "The disaggregation of corporations: selective intervention, high powered incentives and molecular units." STUDI ORGANIZZATIVI.

DECISION SCIENCES INSTITUTE

Innovative Teaching: a Case Study on Teaching Sustainable Supply Chain Management and a Sustainability Mindset

Beate Klingenberg
FOM Hochschule für Oekonomie & Management
Beate.klingenberg@fom.de

Albachiara Boffelli
University of Bergamo
Albachiara.Boffelli@unibg.it

Alice Madonna
University of Bergamo
Alice.Madonna@unibg.it

ABSTRACT

In this research, we propose an exploratory case study for establishing a sustainable supply chain management course ("Sustainable and Global Supply Management", MS in Management Engineering, University of Bergamo, Italy), using a unique, novel combination of pedagogical approaches. To improve learning and encourage the development of the Sustainability Mindset (SM), several exercises aligned with Kolb's experiential learning methodology were implemented. Based on results of the Sustainability Mindset Indicator (SMI®), a measurement of students' attitudes toward the SM, specific activities were deliberately and flexibly chosen. The impacts of the experiential teaching techniques were ascertained by a pre-post analysis using the SMI®.

KEYWORDS: sustainability mindset, sustainable supply chain management, experiential learning

INTRODUCTION

In its various forms, education represents a fundamental building block for society by providing its members with knowledge, skills, values and belief systems. Education shapes what a society is capable of and how younger generations perceive the world. Quoting the Archbishop of York, Eleanor Roosevelt stated in 1930 that "the true purpose of education is to produce citizens" (Roosevelt, 1930). With that, education should reflect the Zeitgeist of its period. Contemplating said Zeitgeist of the 21st century produces a substantial list of issues that our globalized world is facing, one of them being that with continuous population growth the planet earth is not capable of supporting a lifestyle such as found in developed nations for all earthly citizen. In other words, the realization that sustainability is no longer just a trendy word for people with a green mind, but a necessity for humankind that requires a review and revision of values, in other words, a paradigm shift (Davis-Peccour et al., 2017). As Yanarella et al. (2009) state: "In an age of mounting finite resource scarcities, rapid climate change, and continuing global population growth, combined with the growing clamor for Western-style economic development, the sustainability movement is not going to go away" (p. 296).
Since then, new frameworks emerged: the United Nations (UN) formulated the Agenda 2030 (UN, 2015) with seventeen concrete goals, the Sustainable Development Goals (SDGs), as a

guidance for organizations and individuals towards sustainability. Polman and Winston (2021) propose that businesses need to become “net positive”, meaning that they contribute more to the world than they take, while Rimanoczy and Llamazares (2021) promote a needed paradigm shift away from current economic models towards a new mindset of taking care of the planet and society, the Sustainability Mindset (SM) (Rimanoczy, 2021a). The SM evolved from research into motivations of leaders to turn their business into a sustainable one and became a framework for the intentional development of a mindset for sustainability (Rimanoczy, 2010 and 2021b) in educational and training contexts.

Lukman et al., (2021) discussed that various pedagogical approaches are used to teach sustainability, often in combination, such as traditional lectures, case studies, projects, gaming and more. However, the impact of such approaches on learning a “new mindset” is less researched.

Therefore, this paper explores how intentional teaching for sustainability can be assessed, using the case study of a sustainable supply chain management course. The framework of the Sustainability Mindset was used, together with a new assessment tool, the Sustainability Mindset Indicator (SMI®). The SMI® maps an individual’s journey towards such a mindset, allows to identify targeted interventions to develop certain aspects of the SM, and to evaluate pre-post outcomes of these teaching interventions. Furthermore, this research is rooted in Kolb´s experiential learning theory (Kolb 1976, 1984): it challenges traditional learning and teaching methods and suggests that learning happens in a four-phase circular model of different forms of experiencing that create different levels of knowledge. This case study combines the pedagogical options of the SMI® to dynamically address different phases of Kolb´s learning cycle, and to assess pre-post outcomes. The research question is therefore: *How can the learning cycle be dynamically adapted, and the outcomes of the pedagogical interventions be assessed?* While exploratory, it showcases an innovative and not yet reported way of combining pedagogical theory and resources to teach sustainability and assessing teaching outcomes.

The paper Is structured as follows: in the next section, the theoretical frameworks of Kolb´s Experiential Learning Theory and the Sustainability Mindset as well as the SMI® are introduced. A description of the case follows: the course design and the adaptation of the learning cycle are explained. Finally, results and first statistical testing of a sample mindset principles are discussed, using data of the SMI® assessment.

THEORETICAL FRAMEWORKS

Experiential Learning Theory

Kolb introduced experiential learning theory more than forty years ago (Kolb, 1976,1984). A recent meta-analysis of research into the importance and effectiveness of experiential learning (Burch et al., 2019) found that students exposed to this method experienced superior learning outcomes. Narrowing the application to the context of teaching sustainability, Savage et al. (2015) report that students in a sustainability leadership course were excited about experiential learning methods. Particularly, exercises for personal development helped them to gain considerable competences. Sulkowski et al. (2020) used the creation of sustainability reports as a successful, experiential tool to enhance student learning. Dieleman and Huisingh (2006) make the case that experiential gaming plays an important role in building knowledge, skills, and attitudes relevant to capable change agents that support the paradigm shift and transition towards sustainability.

Kolb’s experiential learning model is rooted in neurosciences: according to Zull (2002), different parts of the brain support different types of learning, which are sensing, remembering, theorizing, and acting. Kolb picks up on these learning types and describes it as a continuous cycle of four phases: experiencing (related to sensing), reflecting (related to remembering), thinking (related to theorizing), and acting (Kolb and Kolb, 2018). The cycle starts with experiencing, the assimilative sensing function, resulting in assimilative knowledge. This is also described as sense-making of experiences through apprehension intention by Dieleman and Huisingh (2006). Kolb & Kolb (2018) point out that concrete experiences (vs abstract theories), surprising or provoking aspects particularly enhance letting go of conservative propositions and automatic "everyday" reactions. Put differently, it is not routine experiencing but what James (1977) called "pure" experiences or "shocks" that interrupt normality and trigger reflection that can lead to learning and behavioral changes. The second step is reflective observation or remembering (Kolb & Kolb, 2018). During this step, accommodative knowledge through comprehension intention (Dieleman & Huisingh, 2006) is created and adapted to knowledge gained through experience to new contexts. The result is convergent knowledge through abstract conceptualization (Dieleman & Huisingh, 2006). This thinking or theorizing phase comprises interpretation and analysis, leading to symbolic representation: information is turned into knowledge, ready to be applied within the original context (Dieleman & Huisingh, 2006). Lastly, active experimentation (apprehension extension) leads to divergent knowledge. This is the acting or experimentation phase during which the learner adapts the knowledge to new contexts (Kolb & Kolb, 2018). This active experimentation leads to new, concrete experiences that restart the cycle.

The Sustainability Mindset and the Sustainability Mindset Indicator

The Sustainability Mindset framework was introduced through research by Rimanoczy (2010, 2021b), that took the following perspective: Cronin (2017) coined the term "Age of the Anthropocene", recognizing the ever-more visible human impact on the complex ecological planetary systems. Although faced with a multitude of global, ecological and social crises, and in view of a long history of early warnings (see e.g., Carson, 1962; Meadows et al., 1972; Anderson et al., 2016), surprisingly few significant changes to business models or ways of living have been made. However, in this seemingly deep-rooted complacency, some leaders are sticking out and acting, converting their businesses to more sustainable ones. Taking the angle of Appreciative Inquiry (Cooperrider et al., 2008) and Positive Psychology (Seligman & Csikszentmihalyi, 2014), Rimanoczy's research intended to learn from these leaders' successes, with the objective to develop a framework for higher education to intentionally develop environmentally and socially responsible global citizen (Rimanoczy, 2010). The research asked for which information changed leaders' behavior, how they analyzed it and what motivated unusual steps to change their businesses. This work's outcome is a sketch of how a mindset for sustainability could look like: a way of thinking and being, because of understanding of the ecosystem, developing social sensitivity, and an introspective view on purpose and personal values. The Sustainability Mindset comprises four content areas: ecological worldview, systems thinking, spiritual and emotional intelligence. Each content area is furthermore divided into a total of twelve subtopics, the so-called Sustainability Mindset Principles (SMPs), see Table 1. Each of the principles, in turn, comprise a cognitive, behavioral and affective dimension, as a combination of knowing of *and* being compassionate about something will support behavioral changes.

Table 1: The Sustainability Mindset and its twelve principles (Rimanoczy, 2021b)

Content Area	Principles
Ecological World View	Ecoliteracy, My Contribution
Systems Thinking	Long-term thinking, Both-And, Flow in Cycles, Interconnectedness
Emotional Intelligence	Self-Awareness, Reflection, Creative Innovation

Spiritual Intelligence	Oneness with Nature, Mindfulness, Purpose

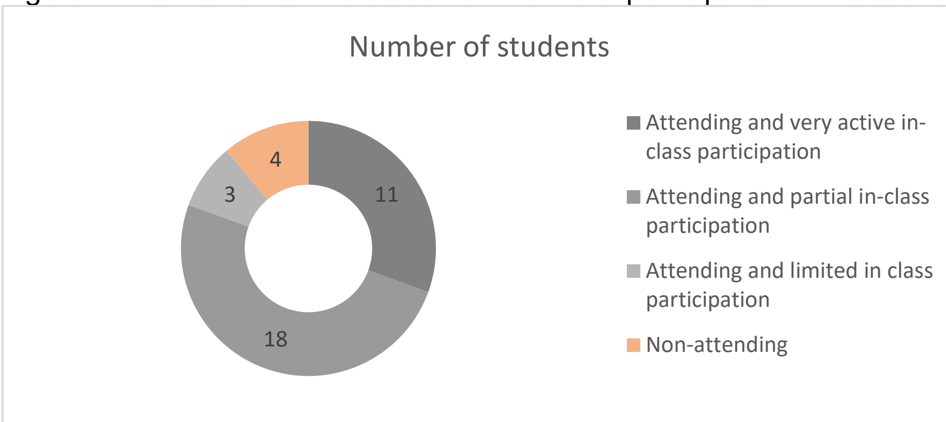
As a next step to make the Sustainability Mindset Principles actionable, an online-based assessment tool was developed that maps where an individual stands regarding the development of the SMPs (Rimanoczy and Klingenberg, 2021). Combining Johnson's (1992) framework of polarities and Kegan's (1994) framework of stages of human development the Sustainability Mindset Indicator (SMI®) was created as a personal development tool. It comprises thirty-six contrary statements, capturing the development of the Sustainability Mindset (Points towards a Sustainability Mindset, PTSM) compared to a conventional mindset (Points towards a Conventional Mindset, PTCM) of the cognitive, behavioral and affective dimensions of each of the twelve principles (Rimanoczy and Klingenberg, 2021). Furthermore, The questionnaire allows to answer "Neither" if a participant cannot identify with either of the statements. The participant is provided an approximately thirty-page personal development report, building again upon the concepts of Positive Psychology (Seligman & Csikszentmihalyi, 2014) and Appreciative Inquiry (Cooperrider et al., 2008). This report elaborates on the principles and provides suggestions and guidance for reflection and experimentation. As the instrument intends to be used in educational settings, instructors also receive a report, which is anonymous and aggregates the data of the individual participants, showing how the Sustainability Mindset is developed across the group. This report allows for identifying weaker developed principles as well as which of the dimensions (cognitive, behavioral or affective) is potentially lacking (Rimanoczy and Klingenberg, 2021). Leveraging this information, educators can then identify specific interventions or exercises that encourage the development these principles with various experiential learning approaches, provided in a set of SMP resource books. (Rimanoczy, 2022). The SMI® furthermore allows for a pre-post analysis: participants can re-take the assessment tool, and the instructor receives a comparative report. It is important to note that this report does not allow for cause-and-effect analysis of interventions taken and observed mindset changes, as the participants are not isolated from other experiences in their everyday lives. However, it does provide a picture of personal development within the group during the period of e.g., a course or other teaching context during which the SMI® was used (Rimanoczy and Klingenberg, 2021).

THE TEACHING CASE

Course Design

This paper is based on a Sustainable Supply Chain Management course that is part of the second year of the Master´s degree program in Management Engineering of the University of Bergamo (Italy). The course recognizes to the students 6 ECTS upon completion of the final exam that can be taken in two modes: as attending student, which means following 48 hours of in-class activities, delivering all the intermediate assignments, and participating in two group projects; or as non-attending student, which means taking an oral exam by preparing all the material provided in-class plus additional readings. The course was delivered for the first time in October-December 2022, involving three educators in the teaching activities and evaluation of the students. All the attending students were also assessed in terms of their proactiveness during the lectures. In total, 37 students took part in the course, and Figure 1 reports their distribution in terms of participation.

Figure 1: Distribution of the students in terms of participation in the course



The course design in terms of general contents and modules was initiated by one of the researchers, with the support and feedback of other colleagues, almost one year before the course start. This period was needed to identify potential tools, educational instruments, and artefacts to be used along the course by participating in teaching sessions at conferences and benchmarking similar courses. Moreover, all the teachers had the chance to train themselves with the tools that were planned to be used along the course: two of the educators received training for the simulation tool that was then used to teach the systemic view behind the circular economy; two other educators received training in the SMPs. Finally, one of the educators was the ideator of the SMI®, allowing to give a complete and rigorous interpretation of the results in real time to the students. This process of continuous learning and improvement enacted by the three educators involved in the course design resulted in an initial course structure that included the following modules:

1) Introduction to sustainability and sustainable supply chain management
2) The internal perspective of the company
3) The Sustainability Mindset Principles
4) The upstream perspective of the supply chain
5) The circular economy
6) Global supply chain management
7) The social side of sustainability

All the modules were designed to include multiple teaching approaches and, when possible, all four phases of Kolb's experiential learning cycle to support the creation of all the four types of knowledge.

Table 2 reports the eight modules together with the tools used, in correspondence with the four phases of the experiential learning cycle, as well as assessment methods used to evaluate the different types of knowledge acquired.

Table 2: Details of the course modules and relation with Kolb's experiential learning cycle

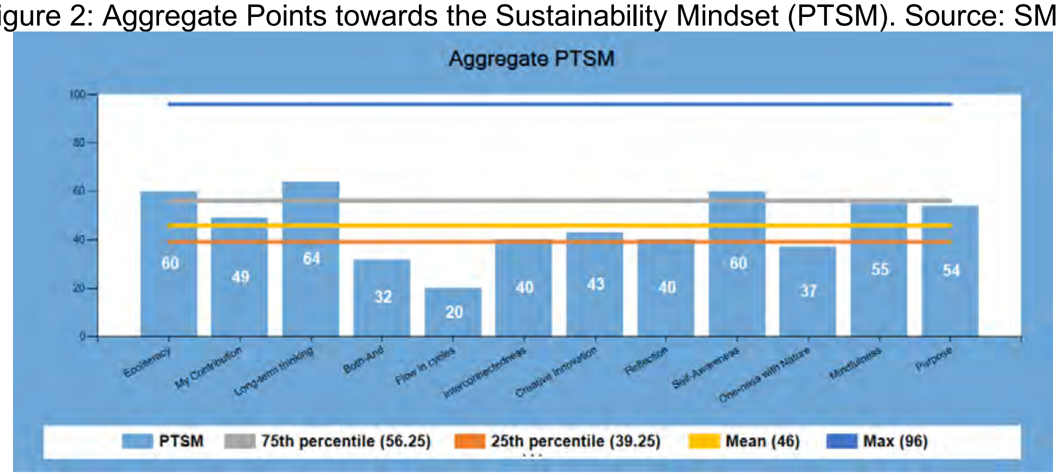
	Experiencing (assimilative knowledge)	Reflecting (accomodative knowledge)	Thinking (convergent knowledge)	Acting (divergent knowledge)
1) Introduc-tion to sustainability and SSCM	Taking the SMI®	Looking at the results from the report	Lectures on sustainability and SSCM	

2) The internal perspective of the company	Reporting recent news on the topics of the lectures	Writing an individual essay including the topic of the module	Lectures on reporting, CSR, SDGs, Change management	Development of a group project work by interacting with a real company
3) The Sustainability Mindset principles	Various experiential exercises	Stop-Reflect approach	Lecture on the Sustainability Mindset principles	Retaking the SMI® and reflecting on the changes
4) The upstream perspective of the supply chain	Reporting recent news on the topics of the lectures	Writing an individual essay including the topic of the module	Lectures on supply chain design, traceability, codes of conduct	Development of a group project work by interacting with a real company
5) The circular economy	Simulation on circular economy	Writing a group report explaining the results obtained from the simulation	Lectures on circular economy	Development of a group project work by interacting with a real company
6) Global supply chain management	Reporting recent news on the topics of the lectures	Writing an individual essay including the topic of the module	Lecture on global supply chain management	In-class debate on why a global supply chain can be considered sustainable or not
7) The social side of sustainability	Showing videos of modern slavery cases (e.g., Shein, Leicester)	Writing an individual essay including the topic of the module	Seminar on social sustainability and modern slavery	
Evaluation of the acquired knowledge	Results of the circular economy simulation	Group report of the circular economy simulation; Individual essay (to include all the modules topics)	In-class participation (as a bonus)	Group project work with companies

Dynamic Adaption and Learning Cycles
Table 2 summarized the course scaffolding. This paragraph elaborates how the content was dynamically adapted based on outcomes of the SMI® as well as interim student feedback. The SMI® was the starting point of the learning cycles in form of a concrete experience to initiate assimilative learning, expecting the thirty-six contrary statements to be a “pure” experience as

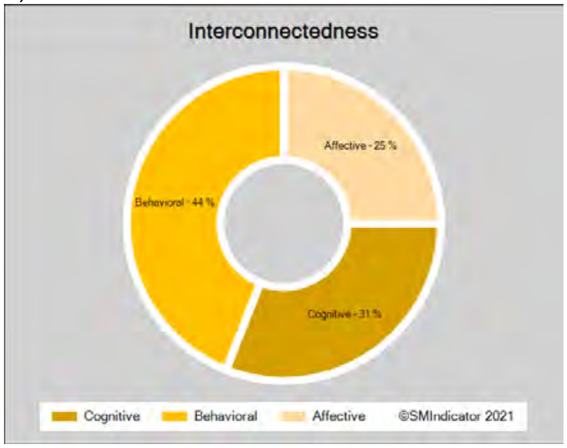
described by James (1977) as an essential beginning. Identifying within these statements the one that reflects a person´s current state of mind is a deeply introspective activity, which then also links to the SMP “Reflection”, in other words, this first activity allowed the participants to immediately experience one of the principles of a Sustainability Mindset and simultaneously, move the students from assimilative to reflective learning.
The second purpose of using the SMI® was to provide the instructors with a snapshot of the mindset of the group at this point of time. Figure 2 is taken from the aggregate instructor report, summarizing the scores of this first administration of the SMI. Shown are the aggregate points towards the Sustainability Mindset (PTSM).

Figure 2: Aggregate Points towards the Sustainability Mindset (PTSM). Source: SMI®.



The maximum PTSM is approximately 100 (the actual number is based on a proprietary algorithm), while the mean PTSM is 46. Weakly developed are: Both-And thinking, Flow in cycles, Interconnectedness, Reflection and One-ness in with Nature.
Furthermore, the aggregate report provides the distribution of cognitive, behavioral and affective dimensions for each of the principles. Taking the example of Interconnectedness, Figure 3 shows the relative distribution of scores for these three dimensions, aggregated for this group.

Figure 3: Cognitive, behavioral and affective dimensions of Interconnectedness



This group expresses the behavioral dimension of Interconnectedness stronger than the other two, with the affective dimension being the weakest.
Based on the combined information, the instructors could identify how to adapt the planned curriculum to address certain aspects. In this paper, examples for Interconnectedness, One-ness with Nature and Reflection are given. Several other interventions specific to the development of the other, weaker SMPs were also introduced to the course, however, the detailed evaluation of their impact is outside of the scope of this paper.
Table 3 summarizes the relative percentages of the dimensions for all three of these principles at the beginning of the course, as well as at the end.

Table 3: Relative percentages of scores at the beginning and end of the course

Principle	Cognitive pre/post	Behavioral pre/post	Affective pre/post
Interconnectedness	31%/31%	44%/38%	25%/31%
One-ness with Nature	31%/35%	16%/23%	53%/42%
Reflection	33%/39%	38%/27%	28%/35%

Examples of interventions to develop Interconnectedness

1) Dragon and Protector (source unknown, cognitive dimension): All participants mix in an open space. Each participant picks a “dragon”, meaning a person to be as far away from as possible and a “protector”, a person to be close to. Who has which role is not disclosed. The exercise generally leads to show-casing systems dynamics: a continuous movement of participants ensues, as they try to move away or get close to a certain person. This exercise again triggered a learning cycle. Reflective observation that followed using a “Stop-Reflect” (see explanation below), allowed students to move the assimilative knowledge of the experience into the realm of accommodative knowledge by the added understanding of systems dynamics. The intervention was aimed at developing the cognitive dimension of Interconnectedness.
2) Coffee Cup (Rimanoczy, 2022, cognitive dimension): Participants are asked to list all people or entities involved in their having a cup of coffee (or beverage of choice). Participants often

start with just a hand-full, sometimes just themselves ("I made my coffee."), however, they soon discover that the list is rather an endless, interconnected system. This exercise adds to the previous by further exploring the systems perspective. From an abstract conceptualization of a system (convergent knowledge), it also represents an active experimentation in new contexts (diverging knowledge). Again, the objective was developing the cognitive dimension of Interconnectedness.

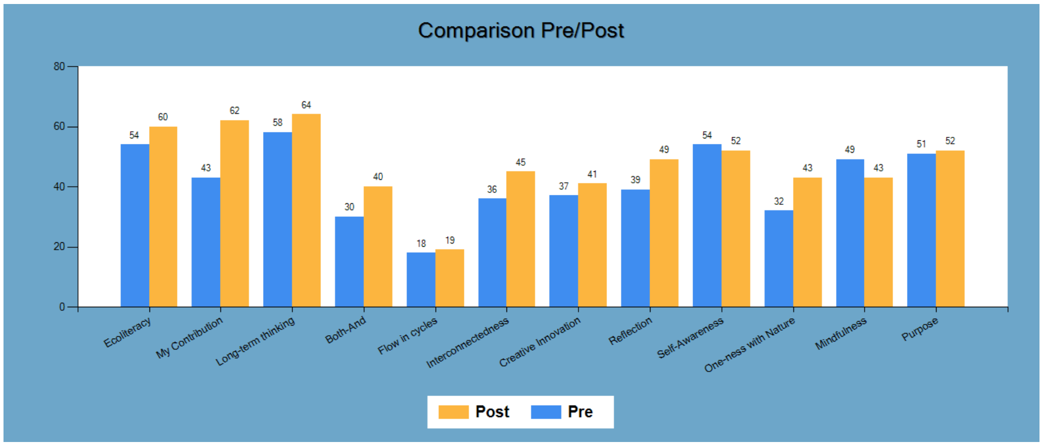
3) My accomplishment (Rimanoczy, 2022, affective dimension): Participants choose a recent personal accomplishment, and then are asked to reflect upon if others than themselves contributed to this accomplishment. Like the previous exercise, the reflection can trigger the realization that even if an achievement is perceived as a personal accomplishment, this is hardly the case, as others generally contribute, even though indirectly. This again triggers reflection on the concept of being interconnected, or part of a system, but approaches it from an affective angle, as accomplishments involve a considerate emotional aspect, such as pride, the feeling of being able to succeed and more. Here, the affective dimension of Interconnectedness was targeted.

Examples of interventions to develop One-ness with Nature.
The course participants were encouraged to spend one hour in nature (Rimanoczy, 2022), without any distractions, such as a mobile device, notepad, book, etc. The intent is to experience the surrounding "unfiltered" and directly, which has the potential of being another "pure" experience (James, 1977), depending on prior habits. The students had the option to reflect upon their experiences in an essay, and to engage in dialogue with each other during a course session. Depending on personal background, this exercise triggers any of the four parts of the learning cycle. It specifically addresses the cognitive side of this principle, as it may create higher awareness of what nature is and how humans experience it.

Examples of interventions to develop Reflection.
Reflective practices were strongly featured throughout the course, using a method called "Stop-Reflect" (Rimanoczy and Turner, 2012; Turner; 2013): students were asked to take approximately two to three minutes to quietly ponder a discussion topic, exercise etc., and then take notes, followed by dialogue in small groups or the plenum, encouraging deeper introspection and reflective (accommodating) learning. This method mostly addresses the behavioral dimensions of Reflection, but also the cognitive one.
Students were also asked to write two reflective essays, one mandatory, and one optional.

ASSESSMENT
Figure 4 shows a graph from the pre-post SMI® report, indicating total PTSM score at the beginning (pre, in blue) and at the end of the course (post, in orange), when students took the instrument for a second time. Some of the students were not present for the second administration, hence absolute scores cannot be compared to the ones in Figure 2. Only the pre-scores of those students that participated twice are shown.

Figure 4 Pre-Post Comparison of aggregate PTSM. Source: SMI®.



Most of the scores increased. Table 4 summarizes the overall (i.e., over all three dimensions) percentage change for the three principles discussed before, as well as p-values of statistical tests. The total number of students participating twice where N=29, however, one response for Reflection was incomplete, resulting in N=28 for this principle. These were performed using Chi-square testing in R, analyzing the pre- and post- proportions of PTSM, PTCM, and "Neither" answers. The significance level is α = 0.05. For each of the tests, the following Null- and Alternative Hypothesis are used (Levine et al., 2017):

H_0: The proportion of scores pre- and post are equal.
H_1: The proportion of scores pre-and post are not all equal.

Significant results are indicated with *.

Table 4: Percentage change of scores and significance		
Principle	Percentage change	P-value
Interconnectedness (N=29)	25%	0.2886
One-ness with Nature (N=29)	34%	0.3825
Reflection (N=28)	26%	0.6537

The results show that – although percentage changes of PTSM are at least 25% - the changes are not statistically significant. Considering that N=29 (28 for Reflection) took the instrument twice, the sample size may be too small to observe a change that is statistically significant.

Table 5 summarizes the results of hypothesis testing for each of the dimensions of each principle.

Table 5: Results of hypothesis testing for each dimension			
Principle	Cognitive	Behavioral	Affective
Interconnectedness	0.5935	0.1985	0.0447*
One-ness with Nature	0.2675	0.7045	0.3473
Reflection	0.5520	0.5520	0.3473

When looking at the changes for each of the dimensions, one sees mostly insignificant results, except for the affective dimension of Interconnectedness. Looking back at Table 3, this dimension sees a gain in relative percentages of 6%, which indicates that this dimension gained in scores. Again, the relatively small sample size should be considered as a potential reason for otherwise insignificant results. Secondly, the second test was conducted right after the end of the course. Recollecting the data after some time (e.g. one year) might show different results.

DISCUSSION, CONCLUSIONS AND LIMITATIONS

On the surface, these hypothesis testing results may seem discouraging regarding the objective of measuring the impact of the chosen teaching interventions. However, as with anything in the context of sustainability, its systemic and often paradoxical nature needs to be considered. The course was designed around sustainable supply chain management principles and the SMPs, thus exploring diverse concepts including rationally driven business frameworks, an ecological worldview, systems thinking, and emotional and spiritual intelligence. The SMPs challenge common value propositions to instill a paradigm shift (Rimanoczy and Llamazares, 2021), which is not something one can expect to happen immediately. Using the pedagogical approach of experiential learning (Kolb, 1974), the students thus encountered a non-traditional and very experience-rich classroom they were potentially not used to, potentially taking longer to absorb everything offered.

Looking at the principle of Interconnectedness, there is a significant change in PTSM scores for the affective dimension. While two interventions were specifically geared towards the cognitive dimension, the reflection on one´s own accomplishment focused on the affective one. One should not overextend the interpretation towards a causal relationship; however, the exercise is powerful enough to realize that an individual hardly achieves anything alone. This could support the statistically significant change in the observed scores.

One-ness with Nature is a principle that may seem “easy” in the sense that feeling comfortable in nature works for most people; however, considering that the generally taught neoclassical economic traditions of seeing nature in the role of providing resources, services, and a place for waste, are challenged when trying to arrive at a more ecological view of the natural environment, it may be understandable that a mindset shift is not immediate.

Reflective practices were exercised throughout the course, and the instructors initially observed surprise and maybe also unease about these by the participants. An important next step in evaluating this reaction is the analysis of the material that is the outcome of the reflecting (e.g., the reflective essays, accommodative knowledge) and acting (e.g., group project report, divergent knowledge) knowledge activities (see column two and four of Table 2). Here, the authors expect to find further insides regarding cognitive, behavioral and affective experiences, and potential mindset changes with regards to Reflection. It is in fact one of the limitations of this paper that the analysis of this material, for example, in form of a sentiment analysis, is not yet included. An in-depth analysis of the other nine of the twelve SMPs is another step that will be covered in future research.

Beyond these limitations, the authors want to point out that for themselves, i.e., as instructors, this new and innovative teaching approach also triggered a learning cycle: starting with the preparation and training received, adapting the syllabus, and continuously observing how the course and its participants developed provided a rich experience, source of personal growth and continuous improvement of the course.

In summary, this case study showcases, to the best knowledge of the authors, a new teaching approach by combining Kolb´s learning cycle theory with the Sustainability Mindset Principles and a new teaching tool, the SMI®. The early results show that the latter tool allows to customize the curriculum according to learning needs, and that – at least on non-statistical levels – changes in the Sustainability Mindset appear to be initiated. It would be interesting to assess on a longer period of time if these changes will be maintained, increased or abandoned, thus assessing the long-term effect of the mindset-changing ad-hoc designed activities and contents. The authors believe that the combination of these methods created an inspirational learning environment that supports learning about sustainability, with all its complexities and paradoxes. This case study may also serve to encourage other educators to try this approach for their teaching contexts.

References

Anderson, T. R., Hawkins, E., & Jones, P. D. (2016). CO_2, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models. *Endeavour*, *40*(3), 178-187.

Burch, G.F., Giambatista, R., Batchelor, J.H., Bucrh, J.J., Duane Hoover, J., & Heller N.A. (2019). A Meta-Analysis of the relationship between experiential learning and learning outcomes. *Decision Sciences Journal of Innovative Education*, *17*(3), 239 – 273.

Davis-Peccourd, J., Stone, P., & Tovey, C. (2017). Fighting sustainability intertia. *Stanford Social Innovation Review*, *15*(3), 59-60.

Dieleman, H. & Huisingh, D. (2006). Games by which to learn and teach about sustainable development: exploring the relevance of games and experiential learning for sustainability. *Journal of Cleaner Production*, 14, 837-347.

Carson, R. (1962; 2002*). Silent spring*. Houghton Mifflin, Boston.

Cooperrider, D., Whitney, D. D., & Stavros, J. (2008). *The appreciative inquiry handbook: For leaders of change.* Berrett-Koehler Publishers, San Francisco.

Cronin, M. (2017). *Eco-translation: Translation and Ecology in the Age of the Anthropocene*. Taylor & Francis, London.

James, W. (1977). Percept and concept: the import of concepts. In: McDermott, J. (Ed.), *The writings of William James*. University of Chicago Press, Chicago.

Johnson, B. (1992). *Polarity management: Identifying and managing unsolvable problems*. Human Resource Development Press, Amherst.

Kegan, R. (1994). *In over our heads: the mental demands of modern life*. Harvard University Press, Cambridge.

Kolb, D.A. (1976). Management and the Learning Process. *California Management Review*, *18*(3), 21-31.

Kolb, D. (1984). *Experiential learning, experiences as the source of learning and development*. Prentice Hall, Englewood Cliffs.

Kolb, A., & Kolb, D. (2018). Eight important things to know about the Experiential Learning Cycles. *Australian Educational Leader*, *40*(3), 8-14.

Levine, D.M., Stephan, D.F., & Szabat, K.A. (2017). Statistics for managers using Microsoft Excel®, 8th edition, Pearson, USA.

Lukman, R.K., Ohmahne, V., el Sheikh, L.T., & Glavič, P. (2021). Integrating Sustainability into Logistics Oriented Education in Europe. *Sustainability*, *13*, 1667.

Meadows, D., Randers, J., & Meadows, D. (1972). The Limits to Growth. In: Robin, L., Sörlin, S., & Warde, P. (Eds.). *The Future of nature,* 101-116. Yale University Press; New Haven.

Polman, P., & Winston, A. (2021). *Net positive – how courageous companies thrive by giving more than they take*. Harvard Business Review Press, Brighton.

Rimanoczy, I. (2010). *Business leaders committing to and fostering sustainability initiatives*. Teachers College, Columbia University. New York.

Rimanoczy, I. (2021a). Anthropocene and the call for leaders with a different mindset", in: Ritz, A., & Rimanoczy, I. (Eds). *Sustainability Mindset and Transformative Leadership: a multidisciplinary perspective,* Palgrave Macmillian, Cham.

Rimanoczy, I. (2021b). *The sustainability mindset principles: A guide to develop a mindset for a better world*. Routledge, Abbingdon.

Rimanoczy I. (2022). *The Resource Workbook for Educators, Vol. 1-4*, Self-published, Amazon.com.

Rimanoczy, I., & Klingenberg, B. (2021). The Sustainability Mindset Indicator: A Personal Development Tool. *Journal of Management for Global Sustainability*, *9*(1), 43–79, https://doi.org/10.13185/jm2021.09103.

Rimanoczy, I., & Llamazares, A. M. (2021). Twelve Principles to Guide a Long-Overdue Paradigm Shift. *Journal of Management, Spirituality & Religion*, *18*(6), 54–76. https://doi.org/10.51327/jkki4753.

Rimanoczy, I., & Turner, E. (2012). *Action reflection learning: Solving real business problems by connecting learning with earning,* Hachette, 91-97.

Roosevelt, E. (1930). Good citizenship: the purpose of education. *Pictorial review,* 31, April, 4, 94-97.

Savage, E., Tapics, T., Evarts, J., Wilson, J.,& Tirone, S. (2015). Experiential learning for sustainability leadership in higher education. I*nternational Journal of Sustainability in Higher Education, 16*(5), 692-705.

Seligman, M. E.,& Csikszentmihalyi, M. (2014). Positive psychology: an introduction. In: Csikszentmihalyi, M. *Flow and the foundations of positive psychology: the collected words of Mihaly Csikszentmihalyi*, 279-298, Springer Verlag, Dordrecht.

Turner, E. (2013). *Gentle Interventions for team coaching: Little things that make a big difference.* Fort Lauderdale, LIM LLC, 130.

United Nations (2015). *Transforming our world: The 2030 agenda for sustainable development*. A/RES/70/1. https://sustainabledevelopment.un.org/post2015/transformingourworld/publication, retrieved March 25, 2019.

United Nations (2021), *Sustainable development report 2021 – The decade of action for the Sustainable Development Goals*, Cambridge University Press, [online] Available at < https://www.cambridge.org/core/books/sustainable-development-report-2021/2843BDD9D08CDD80E6875016110EFDAE> [Accessed April 13, 2022].

Yanarella, E. J., Levine, R. S., & Lancaster, R. W. (2009). Green versus sustainability: From semantics to enlightenment. *Sustainability,* 2(5),296-302.

Zull, J. (2002). *The art of changing the brain*, Stylus, Sterling, VA.

DECISION SCIENCES INSTITUTE

Investigating Personality Types, Attachment Styles, and Behavioral Tendencies

Andrew Hyunwoo Chang
PwC Consulting Solution
andrewwchangg@gmail.com

Sung Hee "Sunny" Park
University of South Carolina
sunny.park@moore.sc.edu

Kealy Carter
University of South Carolina
kealy.carter@moore.sc.edu

ABSTRACT

This research aims to compare the personality type and attachment styles in two different cultural environments, to identify any potential relationships between these variables and behavioral tendencies in the workplace. Two versions of a survey were administered to students at a public university in the U.S. and two public universities in South Korea, respectively. A total of 314 usable responses were used in various analyses. The results demonstrated relationships between these variables are significantly different in different cultures. These findings suggest the continued relevance and value of understanding cultural and behavioral differences through the lens of personality and attachment style.

KEYWORDS: Personality, Attachment Styles, International Comparison

INTRODUCTION

The workplace landscape has long been a subject of scholarly inquiry, with ongoing debates surrounding the prominence of extroverted and introverted individuals within organizational contexts. However, it is imperative to recognize that the efficacy of business practices, the compatibility of personality styles, and the dynamics of interpersonal relationships may vary across diverse work settings. This study aims to contribute to the understanding of these complex dynamics by examining the associations between personality types, attachment styles, and behavioral tendencies within distinct cultural environments.

Within the realm of business education and professional settings, the intricate interplay of personality traits, work preferences, and business-related characteristics significantly influences individuals' performance and outcomes. Moreover, the impact of one's cultural background and developmental experiences on their behaviors and cognitive processes in professional environments warrants careful examination. The conventional dichotomy of introversion and extroversion often serves as a framework for explaining these disparities, despite its inherent limitations in fully capturing the intricacies of individuals' behaviors. In an effort to provide valuable insights, widely employed personality assessments such as the Myers-Briggs and Enneagram have gained recognition in academic institutions and workplaces. These tools facilitate the identification of suitable career paths, foster meaningful connections between students and professionals, and underscore the significance of harnessing individuals' strengths within collaborative work settings (Guy-Evans, 2021).

To expand the current body of knowledge on the subject, this research endeavor endeavors to compare personality types and attachment styles across two distinct cultural contexts. By exploring potential relationships between these variables and behavioral tendencies in the workplace, this study seeks to shed light on the multifaceted nature of human interactions within organizational environments in two distinct cultural contexts. Ultimately, the findings hold implications for organizational psychology, cross-cultural management, and the design of effective human resource practices that accommodate diverse individuals and foster optimal workplace performance.

This paper has the following sections. In the literature review section, we provide a review of personality types (e.g., Introversion, Extroversion, Ambiversion), attachment style, workplace ethics, and social responsibility. We present our research model and hypotheses in the hypotheses/model section. Methods, results, discussion and conclusion sections will be followed. Please note that our research to date is incomplete and we need to do more analyses and update the results, discussion and conclusion sections. A completed version of our research will be available and presented at the DSI 2023 Annual Meeting.

LITERATURE REVIEW

Personality types, such as introversion, extroversion, and ambiversion, significantly influence individuals' behaviors and interactions in social and professional settings. Introverts prefer solitude and reflection, drawing energy from internal experiences rather than socializing. They are often thoughtful, observant individuals who excel in independent tasks. Extroverts, on the other hand, thrive on external stimuli and social engagement. They are outgoing, assertive individuals who excel in group settings and easily establish connections with others. Ambiverts possess a balanced blend of both extroverted and introverted tendencies, adapting their behavior to suit different situations.

Understanding and appreciating the diversity of personality types, including introversion, extroversion, and ambiversion, is vital in creating inclusive work environments that leverage the strengths and preferences of individuals. By recognizing the inherent value that each personality type brings to the table, organizations can promote collaboration, innovation, and overall employee well-being. Furthermore, an awareness of personality types can inform leadership and team dynamics, allowing for more effective communication and tailored approaches to motivation and engagement.

Introversion vs. Extroversion

One aspect of these tests is measuring a person's tendency towards responses to certain relationships, environments, and behaviors; it is commonly referred to as introversion and extroversion. When people first hear introversion and extroversion, they instantly define them as traits. Extroverts are sociable, assertive, and energetic; introverts are reserved, insightful, and passive. In other words, extroverts recharge through social interactions with others, while introverts mainly recharge through spending time with themselves (Eysenck, 2018; Gray, 2017). However, the issue becomes when people formulate presumptions and associate behaviors to how well someone does in their careers, particular environments, and especially within the business world. In many studies, some commonly compared behaviors were happiness, intelligence, work performance, and negative behaviors (McCrae & Costa, 1989; Myers & McCaulley, 1985). Extroverts expressed more positive emotions, unlike introverts who found it neutral (Larsen & Ketelaar, 1991). Introverts were associated with greater levels of intelligence and success in academic settings, even predicting higher grade averages at high school and university levels (Zhang, Zyphur, & Narayanan, 2013). In work settings, extroverts were believed to take on more active roles like sales and teaching, whereas introverts seemed more suited for secluded positions like engineering or accounting (Grant, 2007). Finally, extroverts were more likely to exhibit delinquent behavior in their youth compared to introverts (Caspi et al., 1996).

Many preconceived notions can come with considering yourself as an introvert or extrovert. Characteristics are visible mainly through external interactions, but it also has to do with internal behaviors others cannot see unless shown. Personality can drive certain expectations and motivations; it can impact the satisfaction with a job, or the mannerisms exuded in the workplace. Performance at work is just as important as job security, and one has no immediate control over how to react or handle situations in the workplace. Social relationships and work performance are naturally driven by the people you are surrounded by and the situations you deal with. To some surprise, research studying the relationship between personality and job satisfaction showed that there was indeed a direct relationship between introverts and job satisfaction (Lanaj, Johnson, & Lee, 2016). The relationship was less apparent with extroverts. The reason was how often the two engage with their natural surroundings. While extroverts solely fed off social responses and lacked consideration of themselves, a natural inclination to oneself seemed to show a direct correlation between introverts and work satisfaction. By all means, this research does not apply to all introverts and extroverts, but the general belief was how personality types impacted performance decisions and self-values (Grant & Schwartz, 2011).

With the general comparison of reserved versus outgoing, a more complex definition has to do with one’s tendency for responding to situations. Introverts can generally be seen to react to more negative emotions like stress or fear, whereas extroverts respond more to assertive and sociable settings. How is this seen in the "business world"? In such a fast-paced environment that requires quick thinking, precision, and strong interpersonal skills, is someone likely to perform better with people who are introverted or extroverted? Does it even matter in business? Introvert and extrovert "shaming" is common stereotypical behavior seen when labeling people with certain attributes and behavioral preferences. For example, introverts can be commonly labeled as antisocial while extroverts can be labeled as clingy. This leads to people formulating assumptions and helping to define their "liking" of someone. The barrier between introversion and extroversion is a strange phenomenon that has initiated much controversy over the years. Mostly, within businesses and team-oriented environments, it has become an assumption that introverts are too sensitive while extroverts are too energetic. Hence, many believe that extroverts are more successful than introverts and vice versa with how each side approaches various issues. Over the years, however, these stereotypes have been broken by numerous individuals such as billionaire Bill Gates who is an introvert not fazed by public speaking and artist Adele who is an extrovert admitted to having stage fright.

Ambiversion: Personality at Work

In business, many studies and individuals have displayed behaviors that contradict the assumed work style of varying personalities. A mix of introverted and extroverted traits has developed over the years, also known as ambiversion (Davidson, 2017). Growing up, some embody both traits with a preference towards interacting with people, but a necessity of having to revitalize energy in isolated environments. This new personality type is difficult to distinguish, but a study by Jason Ankeny shows observations and examples of how ambiverts enhance business practices (Weinstein, 2015).

Ambiversion is a borderline term that many do not consider as a trait. Like introverts and extroverts, ambiverts display tendencies that highly depend on situations and communities. A person could be more active and communicate comfortably in one setting while preferring minimal interaction in other environments. Based on that definition, everyone might be considered an ambivert. Carl Jung, who first popularized the terms introversion and extroversion, said that the two differed in how each connected to the internal and external world. Calling extroversion the "normal" personality type, Jung's formulation of introversion arose some debate. Edmund Smith Conklin was a psychologist helping to coin "the ambivert." After reading Jung's writings, Conklin interpreted introversion as a pathological trait (Davidson, 2017). Categorizing people into such clean-cut personality types seemed too ideal and brought attention to how people have dispositions of various personality types. Later, Conklin proposed his own definitions of introversion, extroversion, and the new ambiversion type.

One aspect that distinguished ambiverts from "psychoanalytic or psychological" types were ambiverts' flexibility and adaptability (Weinstein, 2015). He analogized introversion and extroversion as an axis. According to Jung, the relationship between our inner experience (subject) and outer experience (object) is driven by how psychic energy traverses (Davidson, 2017; Weinstein, 2015)3,4 For extroverts, the energy naturally flows towards the object, and they feel comfortable interacting with the outer world; in contrast, introverts energize from self-reflection and observation (Davidson, 2017).

It is important to emphasize that these personality traits do not measure social capacities and abilities to engage in large groups; instead, it is a preference of how one connects and revitalizes themselves to carry forward. These types, however, became nebulous when some started questioning how they exhibit both introverted and extroverted behaviors, especially in different cultural and business settings. Interestingly, Jung argued how pure introverts or extroverts do not exist and acknowledged a third personality type which represented a "normal man;" however, he could not figure out what this was (Davidson, 2017).

Following Jung, many other psychologists, authors, and scientists furthered aspects of these personality types into the third type that Jung was trying to coin. Later, many expanded upon the original two types and looked more into the new ambiversion type. People then started to put introversion and extroversion as opposites on a spectrum. Ambiversion became a hybrid trait, embodying a mix of the two and considered a more adaptable type.

In 2013, Adam Grant, a professor at The Wharton School, tracked various sale representatives at an unnamed software company for three months (Ankeny, 2015). Using a 1-7 scale-based personality assessment, sales representatives who scored between 3.75 - 5.50 (ambiverts) had hourly revenues averaging $154.77 (Ankeny, 2015). Compared to extroverts with $125.19 and introverts with $120.10, ambiverts who scored exactly 4.0 averaged revenues at about $208.34 per hour. As expected, while extroverts were over-dominating conversations during sales pitches, introverts were facing the opposite problem of staying too reserved (Ankeny, 2015). In Han Eysenck's "Dimensions of Personality," Eysenck talks about how extroverts require energizing their minds and bodies which inherently draws out their adventurous, social, and active nature.

On the other hand, in Susan Cain's "Quiet: The Power of Introverts That Can't Stop Talking," many experts debate how introverts are better suited for running businesses due to their greater willingness to hear constructive criticism and their general focused character (Ankeny, 2015). Grant immediately counterargues how both perspectives are not necessarily correct. People who participated in the study showed that many scored in the middle "ambivert" scale (Ankeny, 2015).

There is a comparison about how ambiverts are like bilingual individuals. For example, you can speak English to one person but Spanish to another; the question becomes what your real language is (Ankeny, 2015). Similarly, ambiverts have the advantage of understanding introverts and extroverts equally well. Not only that, but ambiverts who exhibited both extroverted and introverted characteristics proved more successful in general.

Now, ambiverts seem to be that "perfect" model of what all employers covet, so are there any disadvantages? Well, like for introverts and extroverts, the right environment can draw out the best but also the worst performances. Like Grant states, introverts can successfully perform extroverted activities if they find the right niche to grow from, and the same goes for extroverts (Ankeny, 2015). Ambiverts, by all means, are not superior to the other personality types but there are some competitive advantages associated with having a balance of both traits. One problem that ambiverts can run into though is the idea of being overly flexible or unpredictable which can sometimes cause potential setbacks in professional settings (Ankeny, 2015).

Business Ethics

Business work culture plays a crucial role in shaping the preferences of individuals when selecting future companies, firms, and organizations to work for. The ethical practices adopted by businesses have become fundamental to their long-term success. Organizations at all levels, including businesses, firms, corporations, and other entities, place significant emphasis on their work and people culture. Business ethics encompass a wide range of social, environmental, cultural, and legal influences that have the potential to shape or alter individuals' beliefs and behaviors in the workplace. In the present context, upholding high business ethics is not merely about reputation but also about the national image (Sahu, 2021).

Harsha Sahu, an assistant professor at the Management Studies Department of Guru Ghasida Vishwavidyalaya, highlights that business ethics acts as a differentiating factor between companies with high versus low ethical standards. Workplaces characterized by high business ethics and standards tend to exhibit superior decision-making practices, consistent profitability, positive employer-employee relationships, and competitive advantage (Sahu, 2021). Conversely, environments with low business ethics and values can encounter various issues, such as corruption or favoritism, which can give rise to negative work culture, underutilization of employees, and potential legal repercussions (Sahu, 2021). Business ethics serves as an entry point for companies when interacting with other organizations or clients, and it also acts as a magnet for attracting quality employees (Sahu, 2021).

Within organizations, the workforce is composed of individuals from diverse backgrounds, possessing varied technical skillsets and interests. Hence, it becomes imperative to examine how business-minded students with these differences perceive business ethics and culture, considering the influence of various behavioral factors and experiences. Understanding these dynamics can shed light on the complex interplay between individual characteristics and organizational environments, leading to valuable insights for both academia and industry. This study aims to explore and analyze the perceptions of business-minded students regarding business ethics and culture, taking into account the multifaceted influences shaping their attitudes and behaviors in this context

Attachment Style

Attachment style, a construct derived from attachment theory, has gained attention in the study of business ethics and consumer behavior (Scott, 2008). Attachment theory explains how social cognitive representations drive human needs to establish interpersonal relationships (Scott, 2008; Thomson & Johnson, 2012). While attachment theory has traditionally focused on psychological attachments such as caregiving and romantic relationships, its application to consumer attachment styles has gained interest. Consumer attachment styles refer to the ways individuals relate to brands, services, and business-related concepts (Thomson & Johnson, 2012). This article explores the manifestation of attachment theory in the business world and work culture, shedding light on its implications for ethical practices.

Attachment style is characterized by two dimensions: attachment avoidance (AV) and attachment anxiety (AX) (Thomson & Johnson, 2012). Attachment avoidance pertains to fears of personal intimacy, dependence, and disclosure, while attachment anxiety relates to fears of rejection and abandonment. These dimensions give rise to four attachment styles: secure, preoccupied, dismissing, and fearful (Thomson & Johnson, 2012). Secure individuals have positive representations of both themselves and others, while preoccupied individuals have negative representations of themselves but positive representations of others. Dismissing individuals exhibit negative representations of others, and fearful individuals have negative representations of both themselves and others. Figure 1 displays the relationships between attachment avoidance, attachment anxiety, and the four attachment styles.

Figure 1

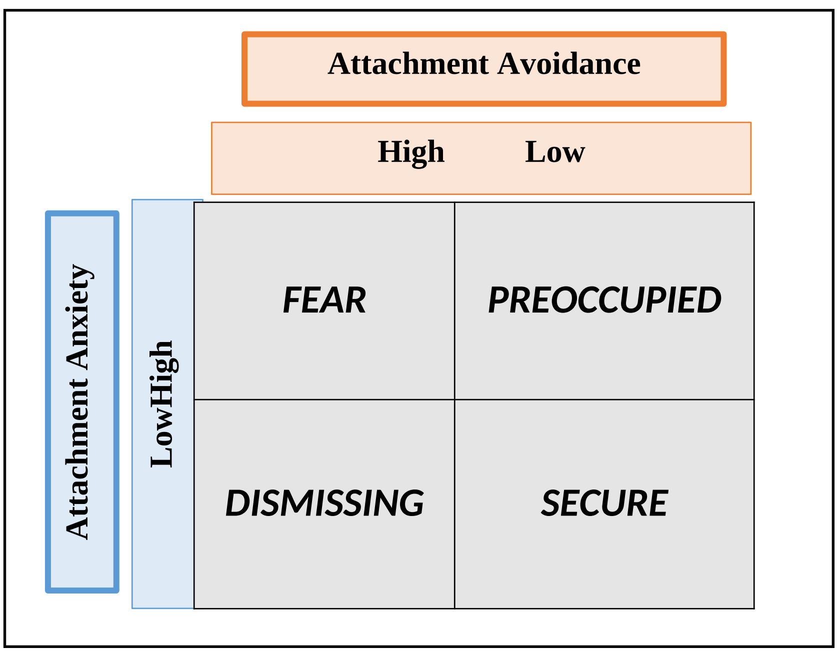


Figure 1: *Relationship between attachment avoidance, attachment anxiety and four attachment styles*

Previous studies have explored the influence of attachment styles on various consumer behavior variables, such as the expression of attitudes and complaints (Thomson & Johnson, 2012). These studies utilized models to assess whether attachment avoidance and anxiety were separate yet interconnected dimensions or completely unidimensional in consumers' self-expression. The results supported the distinction between avoidance and anxiety as distinct dimensions (Thomson & Johnson, 2012). Moreover, secure individuals were found to be less likely to complain or act against brand actions, while fearful individuals demonstrated a higher tendency to engage in such behaviors (Thomson & Johnson, 2012). Further research examined the impact of self-image and loss of benefits on attachment avoidance and anxiety, corroborating their reliability as metrics for assessing individual traits (Thomson & Johnson, 2012). Additionally, investigations into the correlation between attachment styles and personality traits have been conducted to determine their respective influences on business ethical practices.

The examination of attachment style and its implications for business ethics and consumer behavior provides valuable insights into the interplay between individual characteristics and organizational environments. Attachment style has emerged as a relevant construct within the business domain, shaping consumer behavior and influencing ethical practices. By understanding the relationship between attachment style and consumer behavior, businesses can tailor their strategies to effectively engage with consumers and foster ethical business practices.

Workplace Ethics & Social Responsibility

Globalization has inevitably impacted the way we conduct and view business. Globally, we see how virtue ethics and social responsibilities are highly emphasized and valued in business communities. In successful work environments, employee well-being, flexible work schedules, and team transparency are frequent practices seen. Over the years, many debated whether organizations could be independently moral while sustaining a certain level of personal growth for employees (Dawson,2018).

Organizational virtue (OV) and business ethics have proven to carry out an organization's goal and showcase how direct business influences impact individual values (Dawson,2018). Many people have created OV scales to measure organizational virtues in the workplace (Dawson,2018). Some focused on organizational practices and behaviors, while others on moral attentiveness and structural relationships (Dawson,2018).

In the past, there has been an issue with how virtue and standards were implemented in workplaces. In the United Kingdom (U.K.), a survey was conducted to measure business ethics and the relevance of its impact on undergraduate students. Approximately 440 students were selected from a participant poll, and out of the 440 students, 385 graduated from a university course (Dawson,2018). With response rates of about 40%, the goal was for respondents to finish all the OV scales in the survey. The results showed that 32% of the respondents were able to complete all the scales successfully (Dawson,2018).

The appeal is geared towards students and individuals specified in the business field. In order to teach young adults before entering into the real world, this human resource study aims to explain business ethics when working with others. A few hypotheses were formulated in the process such as indications of OV being positively correlated with moral attentiveness (MA) and higher MA leading to more visibility of ethics and social responsibility (Dawson,2018).

Even though there were some positively correlated relationships with moral attentiveness, it showed that there were some negative correlations with OV trust, forgiveness, and relationships. From most business students' perspectives, many agreed that social responsibilities and organizational ethics were vital to workplaces (Dawson,2018). However, many also were uncertain about the priorities of the attention given to business issues versus the students themselves (Dawson,2018). Studies like this have shown the relevance and importance of alleviating pressured work environments, workload stress, valuable work experiences in achieving compatibility between ethics and business (Dawson,2018).
In this article, a similar approach was taken, where personality types and attachment styles were taken into consideration for understanding how students could view certain ethical practices. Based on business students' academic backgrounds, career goals, and social aspects of real-work experiences, the research compares the similarities and differences between two academic and cultural settings.

HYPOTHESES/MODEL

When comparing the similarities and differences between two different cultural settings, the first goal was selecting variables that may impact how subjects behave in professional settings. Since we used students at a public university in the U.S. and two public universities in South Korea as subjects, we need to make sure selecting variables that may impact how students behave in professional settings. For example, culture and family backgrounds, different educational mindsets, the types of work experiences, etc., were all considered when surveying subject in both cultures. Based on that, there were two specific hypotheses formulated.

> H1: Personality types have an impact on the relationship between working environments/cultural settings and behavioral tendencies in the workplace.
>
> H2: Attachment styles have an impact on the relationship between working environments/cultural settings and behavioral tendencies in the workplace.

All factors aforementioned are undeniably relevant to how students perform and learn. Education, for instance, is a big part of the many underlying differences in student educational mindsets, even just between education systems in the United States and Korea. Based on how the education system in Korea is constructed, and from the academic standards that most Korean households uphold, students invest longer hours solely on academics and use that as a driver for success.

On the other hand, education in the United States encourages individuality, where students pursue opportunities that pique their interests and become well-rounded. It results in a disparity between East Asian cultures, most notably Korea, and Western cultures in how students perceive their accomplishments. Students in the United States are more inclined to overestimate their abilities and believe that high self-esteem leads to success, while students in Korea tend to underestimate their abilities and believe only self-improvement leads to success.

It is inevitably a big part of one's thought development process and behavioral patterns. Culture and family backgrounds also play into one's personality and professional demeanor. Although it was not heavily studied, cultural and family influences were also considered.

Work experiences were the main focus of the surveys. As the majority of business students have had past or have current and potential job offers, internships, and other professional opportunities, career experiences were used as the basis for the study. The beginning of a business student's career can vastly differ by individual. Some college students start with side jobs like fast food or academic sectors, while others begin working directly in their respective fields — financial analysts, brand ambassadors, consultants, administrative assistants, etc. In the workforce, one commonality is when students try to fit in with preconceived standards of their work environment. In other words, extroverts try to be introverts and vice versa because they believe certain behaviors or skills are favored. It is like having an "on-off" switch between the personality types.

Personality influences a multitude of factors, not just associated with behaviors. It controls a pattern of actions and emotions that elicit specific responses, especially in unfamiliar settings. Introverts and extroverts may experience similar internal and social anxiety in new work environments, but their behavioral patterns and coping approaches could be different. Introverts are typically more socially withdrawn, while extroverts are more socially dependent in active environments. It is not a matter of how social one is but how often one gains energy and presents themselves in social settings.

However, this relationship between introverts and their behavioral tendencies may differ in professional settings. First impressions play a big part in how employers and other co-workers view an incoming employee. Introverts will feel obligated to engage in social activities and in-person events, thinking that active social engagements will be particularly favored.

This study tests how additional factors, like work experiences and leadership positions, directly impact student behaviors among introverts and extroverts. Based on student presumptions of how successful personality types would be in various professional settings, introverts are more likely have the "on-off" personality switch and expect to be more extroverted/introverted than do extroverts in work environments (Thomson & Johnson, 2012). When placed in work settings, introverts who have more work experience or leadership positions will show more extroverted tendencies while extroverts will maintain their extroverted personality.

Attachment style was another trait variable used for studying business student behaviors and overall thought processes. With attachment theory, the study focused more on the “relationship” aspect of business students and how their emotional connections impacted their ability to interact with others. Attachment theory started with mainly referring to individual social and romantic relationships but gradually evolved into being described for business and marketing. Researchers believed that attachment styles could be used to “predict” business and consumer relationships. Many believe that “attachments” to business partners, brands, service providers, etc., could lead to anti-brand behavioral actions and future relationships.

Specifically, two business hypotheses were studied. The hypothesis was based on anti-brand behavior and the theory that individuals with high anxiety and avoidance — otherwise considered fearful consumers — would be most likely to experience anti-brand reactions. Fearful individuals are more likely to invest more into the business, which results in negative psychological impacts of lower self-esteem or self-reflection. In other words, the higher the attachment anxiety (AX) and attachment avoidance (AV), the greater the investments individuals are likely to make to companies and brands, resulting in those reactions. Results showed that attachment avoidance and anxiety were two separate yet distinct dimensions that clearly showed how individuals with low scores on both were less likely to initiate anti-brand actions like complaining or payback.

Later, the study took the first hypothesis a step further by adding self-image and loss of benefits mediators into the equation. The relationship between avoidance and anxiety was then studied again, and it was proven that the metrics were precise in assessing individuals' responses (Thomson & Johnson, 2012). When looking at self-image and loss of benefits independently, the metrics were less statistically significant, but, in conjunction, were highly correlated to the anti-brand behavioral patterns (Thomson & Johnson, 2012). In conclusion, it was shown that anxiety and avoidance correlate to business actions and relationships. This research was conducted to take a similar approach but combining individualistic and business ethics viewpoints.

The study worked to integrate both the social and business-related attachments that students perceived when thinking on a business level. Using some questions that assessed attachment anxiety and avoidance in the survey, the second hypothesis wanted to investigate how business students felt about their relationships, specifically in the business world. If the first hypothesis was looking more into the personality aspect of students, this wanted to get some insight on how students engage with their environment and go about their relationships in business settings. After analyzing the survey responses, the goal was to see if there was a correlation between personality and attachment theory or if one was more statistically significant.

Attachment anxiety and avoidance both impact the way people build relationships, and the study looked to identify if there were any differences for students in professional work settings. Not only that but does attachment style impact the personality types? While researching attachment styles and looking into studies that show how different attachment levels impact individuals, there was some correlation between anxious or secure individuals and predictions of their behavioral patterns (Thomson & Johnson, 2012). The ways one emotionally connects and approaches relationships could be predicted through these patterns. Later, this then brought up the question of whether attachment styles also predict levels of introversion and extroversion. Were introverts and extroverts more highly anxious, avoidant, secure, or fearful, and could this be used to predict certain behavioral tendencies?

There are four types of attachments which are secure, avoidant, anxious, and disorganized/fearful attachments. Each of these attachments is based on the relationship and different levels of AX and AV. Higher or lower levels of anxiety and avoidance can show signs of a more secure or fearful individual. Introverts are believed to have higher levels of anxiety and avoidance due to their naturally reserved and perceptive personalities. For extroverts, attachment levels may seem lower because their personalities allow them to cope with challenges using their more outgoing and gregarious personality. In business environments, the first steps in meeting new people and integrating into the workplace have a lot to do with how individuals mentally prepare to start a connection. By testing the relationship between attachment styles in introverts and extroverts, the study wanted to see if students displayed certain levels of AX and AV in their respective academic and work settings.

Although there may be more secure individuals in interpersonal and comfortable social settings, this could be the opposite for business students who constantly engage with new people in fast-paced work environments. It is predicted that AX and AV are higher in business students, which could inevitably impact student perceptions of the work culture, their capabilities, and professional relationships. U.S. and Korean education systems and work cultures are vastly different. Unlike the Korean work culture, the American work culture has many more direct lines of communication with its employees. In other words, it is much easier to express opinions, thoughts, and decisions with anyone at all hierarchical levels. Korean work culture is more individualistic with how Koreans handle situations. They are more of the "keep-it-to-yourself" type and find it hard to open up about struggles. They believe figuring problems out alone is easier, even if it takes longer. This may also have to do with the intense hierarchical levels of Korean work culture compared to the U.S. work culture. In Korean work culture, formalities are not only critical but mandatory for thriving in business settings. While incoming employees can be a little more casual in their interactions with other staff, Koreans take formal language very seriously, even when interacting outside of work. These differences and factors can predict how U.S. and Korean students respond to and build relationships. Due to these differing work and social factors, U.S. introverts and extroverts will show higher signs of anxious attachment, while Korean introverts and extroverts will display higher levels of avoidant attachment. In addition, there will be significant differences in attachment styles between introverts and extroverts, both in the U.S. and Korean university student population.

METHODS

Survey Design

The first step of the research process was collecting data from surveys that would explore business student perspectives and their thoughts on how introverts and extroverts successfully incorporate business practices worldwide. Two groups of students were surveyed. The first was students in a large state university in U.S., which broadened the scope of individuals to those who would possibly utilize and experience ethical business practices in the future. Since there are in-state, out-of-state, and international students, it helped diversify the survey with people within and outside the United States.

The second group was business students at two South Korean universities (Korea Tech and CNU). The goal was to compare these two business subgroups from two different environments and cultures, seeing if there was a relationship between them and the business ethical principles they value in various workplaces and settings. There were two versions of this survey: one in English and one in Korean.

Aside from some questions like the student’s university and major that the Korean version asks to list out — since students will be from different universities with possibly different majors — the questions were the same in both versions. Each survey was practically identical, with the only difference being the language and some questions relating to the business curriculum specific to each university. All responses remained anonymous, and the only identifier was the student university email to ensure that only U.S. students responded to the survey.

For the U.S. survey, there were four specific sections. The first section required respondents to answer background questions like the year, major, and hometown to gather general demographics of students at the university. The last question in that section asked whether the students considered themselves introverts or extroverts. The next section was a transition to introversion and extroversion in business. These questions were a little more probing, where the goal was to see if students thought any differently in a business context, giving their views on business introverts and extroverts. The third section asked how business students defined introversion and extroversion. It goes back to the preconceived notions research part of the thesis to see how students defined these personality types.

The final section of the survey asked questions on attachment anxiety and attachment avoidance. Respondents were asked to answer each statement — with no context on the attachment styles — and the extent to which they agreed or disagreed with each one. These questions were given on the Likert scale, where the choices given were strongly disagree, disagree, neutral, agree, and strongly agree. The goal of these questions was to have students rate their agreement to statements. For this section, there was no description of what these questions were measuring but prompted how students felt about their relationships with close others — a broader way of describing attachment avoidance and anxiety.
For the Korean survey, the state questions and email portion were excluded. Sections 2-4 were identical but just in Korean.

Data Collection

The data collection of students started in February 2021 until March 2022. Since then, one year's worth of data was collected resulting in the final sample size of around 253 U.S. students and 131 Korean students. This survey was the first step in initiating the analysis. It was also what took the longest due to different academic seasons where students would be on break or busy seasons at the three universities. All results were then compiled into an excel spreadsheet and organized in the Tableau software.

Due to COVID-19, there was no opportunity to go abroad, making it harder to collect data in person, especially for the Korean cohort. Through friends and DMSB professors, anonymous google surveys were distributed and administered via email. The only compensation given was a gift card to one random student that participated in the survey. No other rewards or compensation were awarded for participation.

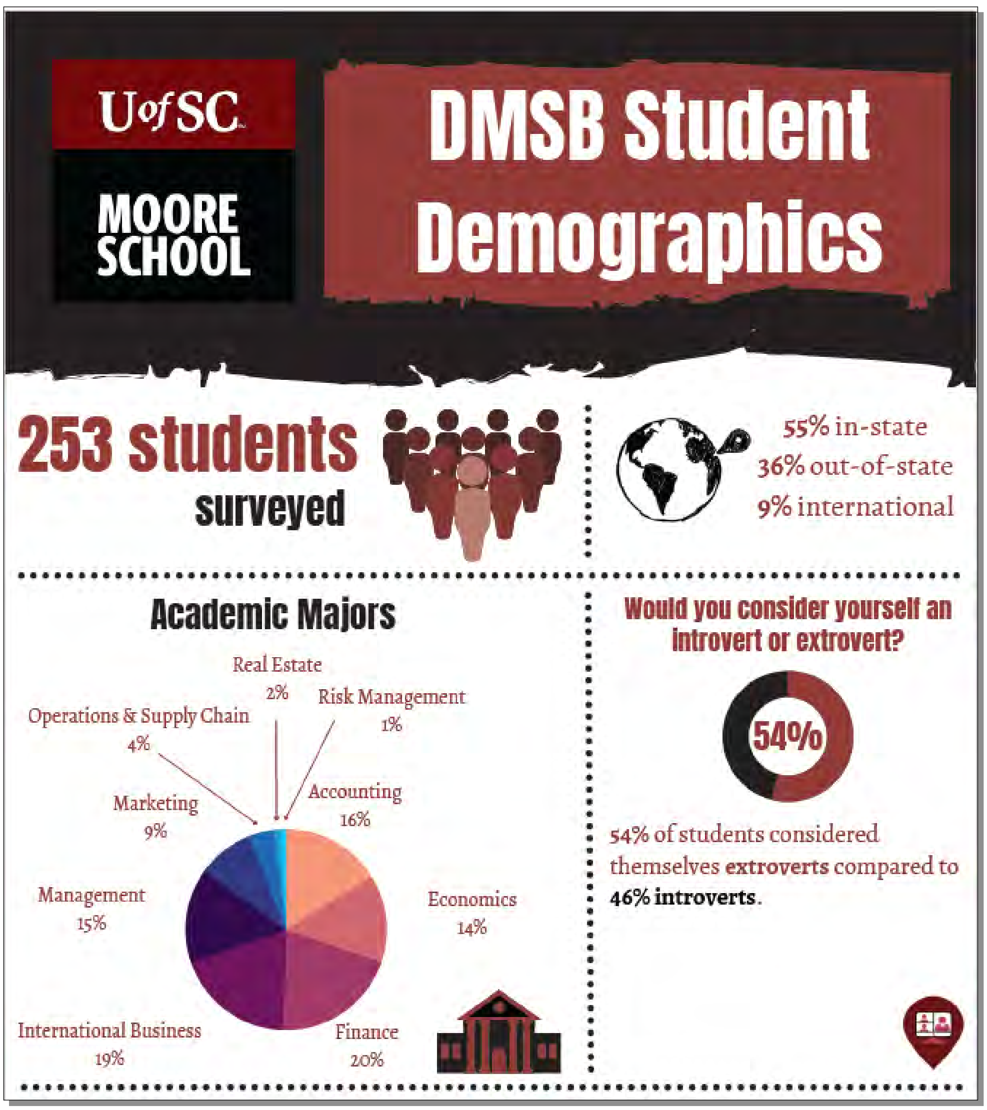
Data Analysis

We are going to utilize a MGSEM to test our hypotheses. As we mentioned, our research to date is incomplete and we need to do more analyses and update the results, discussion and conclusion sections. A completed version of our research will be available and presented at the DSI 2023 Annual Meeting.

RESULTS

The U.S. survey was sent to students at Darla Moore School of Business (DMSB) at the University of South Carolina. For U.S. subjects, the ratio of extroverts and introverts was similar across all majors. As seen in Figure 2 below, the percentage of introverts and extroverts, their demographics, and majors within the student pool was compiled. There were 137 students (54%) who classified themselves as extroverts, while 116 (46%) students classified themselves as introverts. Looking into the demographics helped to better understand the student population being studied.

Figure 2: *Surveyed Darla Moore School of Business (DMSB) Student Demographics*



Respectively, Tables 1 and 2 break down the number of students per major and whether they considered themselves an introvert or an extrovert. This initiated the first step of the analysis to see whether the responses were similar across majors. With 70% comprised of accounting, finance, international business, and management majors, it was interesting to see how the majority of certain majors like finance and international business, were extroverts, and the vast majority of accounting majors were introverts.

Student major(s) at DMSB

Majors	
Accounting	60
Economics	55
Finance	77
International Business	73
Management	58
Marketing	34
Operations and Supply Chain	17
Real Estate	5
Risk Management and Insurance	3

Table 1: *# of students per major*

Darla Moore School of Business (DMSB) Extroverts and Introverts Demographics

Would you consider yourself an introvert or extrovert?	In-state	International	Out-of-state
Extrovert	75	8	54
Introvert	64	15	37

Table 2: *Domestic/international demographics for DMSB students and what personality type — extrovert or introvert — each student classified themselves as*

The survey was also sent to two Korean universities, Korea Tech and CNU, where business students were surveyed. It was much harder to gather many responses from the Korean cohort due to the differing academic seasons and wait times. As a result, the number of students surveyed was relatively low compared to the DMSB student population. Although there were only about 131 students surveyed, information from two universities contributed to the diversity of the results. As seen in Figure 3 below, there was a wide range of majors across the universities ranging from business administration, accounting, and economics to agricultural chemistry, software engineering, and English literature. Most students majored in business administration. Similar to the DMSB student population, 52% of Korean students considered themselves extroverts (compared to 54% of DMSB students), and 48% classified as introverts (compared to 46% of DMSB students). Korean university students were also prompted to answer if they felt more introverted or extroverted in the business context.

DISCUSSION AND CONCLUSIONS

This research analyzed two concepts and studied the relationships between business students and their personality and attachment levels. By sampling two student groups — Darla Moore School of Business (DMSB) students at the University of South Carolina and students at Korea Tech and Chonnam National University (CNU) in South Korea — the study focused on analyzing the relationship between personality types and attachment styles to behavioral tendencies in professional environments.

The first relationship between personality types and behavior found that introverts were more likely to display extroverted characteristics than extroverts with introverted tendencies. However, introverts and extroverts alike in both groups had common responses toward the characteristics, behavioral tendencies, and ethical practices they commonly see or would like to see in business settings. In conclusion, the first hypothesis was partially supported.

With attachment styles, some evidence through the survey proved that attachment anxiety and avoidance had direct relationships to student behaviors. Although the questions were somewhat surface level, the responses from both groups tied to the research of personality types and showed relationships to how students in the U.S. think and react differently than students in South Korea. Overall, attachment anxiety and avoidance were low, implying that many students were secured individuals. However, there was more of a mix between the levels of agreement, making it difficult to establish an exact relationship.

Regardless of the cultural background, work experiences, personality types, and attachment styles, the study concluded that there were relationships between introvert and extrovert personalities and secured, preoccupied, dismissing, and fearful attachment styles. Although business culture is vastly different, business relationships and work environments revolve around various perspectives. It is important to realize that personality types or levels of attachment styles do not define or make someone more successful. Many people often show both signs of personality traits or differing levels of attachment anxiety and avoidance. Different behavioral tendencies may be more favorable in establishing relationships within varying professional fields and cultures. As such, every experience and relationship one encounters only enhances and contributes to one's ever-growing growth and success.

REFERENCES

Ankeny, Jason. 2015. “A WINNING PERSONALITY. (Cover Story).” Entrepreneur 43 (3): 36–41.

Caspi, A., Elder, G. H., & Bem, D. J. (1996). Moving against the world: Life-course patterns of explosive children. Developmental Psychology, 32(4), 604-614.

Davidson, Ian J. “The Ambivert: A Failed Attempt at a Normal Personality.” Journal of the History of the Behavioral Sciences 53, no. 4 (October 2017): 313–31. doi:10.1002/jhbs.21868.

Dawson, David. 2018. “Organisational Virtue, Moral Attentiveness, and the Perceived Role of Ethics and Social Responsibility in Business: The Case of UK HR Practitioners.” Journal of Business Ethics 148 (4): 765–81. doi:10.1007/s10551-015-2987-4.

Eysenck, H. J. (2018). The biological basis of personality. Routledge.

Grant, A. M. (2007). Relational job design and the motivation to make a prosocial difference. Academy of Management Review, 32(2), 393-417.

Grant, A. M., & Schwartz, B. (2011). Too much of a good thing: The challenge and opportunity of the inverted U. Perspectives on Psychological Science, 6(1), 61-76.

Gray, J. A. (2017). The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system. Oxford University Press.

Lanaj, K., Johnson, R. E., & Lee, S. M. (2016). Benefits of transformational behaviors for leaders: A daily investigation of leader behaviors and need fulfillment. Journal of Applied Psychology, 101(2), 237-251.

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. Journal of Personality and Social Psychology, 61(1), 132-140.

McCrae, R. R., & Costa, P. T. (1989). Reinterpreting the Myers-Briggs Type Indicator from the perspective of the five-factor model of personality. Journal of Personality, 57(1), 17-40.

Myers, I. B., & McCaulley, M. H. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Consulting Psychologists Press.

Sahu, Harsha. 2016. “A Descriptive Study on Business Ethics.” CLEAR International Journal of Research in Commerce & Management 7 (9): 35–38.’

Scott LN, Levy KN, Pincus AL. Adult attachment, personality traits, and borderline personality disorder features in young adults. J Pers Disord. 2009 Jun;23(3):258-80. doi: 10.1521/pedi.2009.23.3.258. PMID: 19538081; PMCID: PMC3195524.

Thomson, M., Whelan, J., & Johnson, A. R. (2012). Why brands should fear fearful consumers: How attachment style predicts retaliation. Journal of Consumer Psychology, 22(2), 289–298. https://doi.org/10.1016/j.jcps.2011.04.006

Weinstein, Ian. (2015). Learning and Lawyering across Personality Types. Clinical Law Review, 21(2), 427–453.

Zhang, Z., Zyphur, M. J., & Narayanan, J. (2013). The ties that bind: A social network approach for investigating the mechanisms underlying the psychological success of transformational leadership. Journal of Applied Psychology, 98(2), 264-278.

DECISION SCIENCES INSTITUTE

Investigation of Economic Growth in OECD Countries: A pre vs. post pandemic outlook

Alicia M Macias
Department of Economics
aliciamacias702@gmail.com
Plaster College of Business and Entrepreneurship
Lindenwood University

Gokhan Egilmez, Ph.D.
Department of Management
gokhanegilmez@gmail.com
Plaster College of Business and Entrepreneurship
Lindenwood University

ABSTRACT

This research paper explores the role of COVID-19 in the OECD economies by examining and comparing commonly used indicators from 2016 to 2021. Key economic indicators were utilized to assess economic growth by using statistical analysis and visualization. Statistical tests indicated that inflation and total factor productivity worsened significantly during and after pandemic. It was observed that the unemployment rate reached its highest level since the global financial crisis of 2008-2009 period. Conversely, research and development funding was increased, while governments worldwide increased budgetary allocations to support unemployed workers, enhance healthcare facilities, and aid their citizens.

KEYWORDS**:** economic growth, OECD, COVID 19, Pandemic, statistical analysis

INTRODUCTION

On January 2020, The World Health Organization declared COVID-19 a global emergency, a few months later it was declared a high level of health emergency. Many countries implemented a public policy in order contain the virus, so it does not spread, such as quarantines and people working from home. These public policies affected the $90 trillion global economy (Congressional Research Service, 2021). The was a decline were the lowest was -3.1 on GDP Growth (The World Bank, 2021) between the years 2016 and 2021. In this research paper we investigate critical measures of economic growth in order to compare pre and post pandemic economic growth on select OECD countries.

This research paper explores the role of COVID-19 in the economy by examining and comparing economic trends from 2016 to 2021 in OECD countries (Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and The United States), it examinates OECD economic indicators due to the data availability and quality founded as this organization possess reliable and comprehensive economic data that are available for research. Additionally, as English is one of the most used languages for academic research publications making easier the access and understanding to them in a commonly understood language. OECD is a highly recognize organization and the countries that are part of it, holds a high global economic significance, having a standardized report format that made easier to compare data sets and data analysis across the 38 countries that are part of it.

Key economic indicators were utilized to assess the growth of the economy by using statistical data visualization. It was observed that the unemployment rate reached its highest level since the global financial crisis of 2008- 2009 during this period due to stringent restrictions imposed by governments and employers to curb the spread of the virus. The closure of factories disrupted the supply chain, resulting in higher prices for goods and services, which were limited in availability, and primarily accessible to individuals with greater purchasing power. Conversely, research and development organizations augmented their funding for further research and technological advancements, while governments worldwide increased budgetary allocations to support unemployed workers, enhance healthcare facilities, and aid their citizens. Based on the findings of this study, the following recommendations can be made, the diversification of supply chains, technology investments and the promotion of a sustainable development.

LITERATURE REVIEW

Economic growth is defined as the increase in the production of goods and services per head of population over a period of time (Oxford, 2006). Economic Growth is significance because measure the economy health in a country, therefore, is an indicator of wealth. It measures the increased of the real GDP and the increase of national product that measures the constant prices of the economy (Denison 1962). The four major determinants of economic growth are human resources, capital formation, and technology (Hicks, 2021), these factors are needed for an efficient market; otherwise, the supplier may face higher costs of production. (Castaneda, 2021). When the economy growth increases, countries are able to increase state capacity and the supply of public goods, as a result it creates wealth that improve the wellbeing of the population by increasing their living standards (Sen, 2021).

Economy indicators allow economist to analyze the global economy activity. Every country produces goods and services; therefore, it measures the changes on the structure of countries, and it helps to identify their growth or contractions. (The World Bank, 2023). In order, to determinate whether the economy is growing or declining, economists uses measures of macroeconomic performance such as the Gross Domestic Product (Helpman, 2004), Consumption (The World Bank, 2023), Government Expenditures (Boldeanu, Constantinescu, 2015), International Trade. The System of National Accounts provides an integrated system of accounts that helps in the comparison of the important economy activity, as the Balance of Payments helps to monitor a country's international transfers, and the Government Finance Statistics observe government income and expenses (The World Bank, 2023). GDP is a widely used indicator, therefore if there an increase of growth rate there should be an increase on each component that has an effect on GDP. (Boldeanu, Constantinescu, 2015). If the constant prices changes, it will be reflected in the change of GDP. Updates about economic indicators data occur every July and December, in some case data can be updated more often when countries updated their data monthly or quarterly in order to introduce or revise their data. The Atlas method is used to classify countries depending on their income level. This method classifies economies into four categories for analytical purposes: low-income, lower-middle-income, upper-middle-income, and high-income (The World Bank, 2023).

Capital Goods, Labor Force, Technology and Human Capital (Boldeanu, Constantinescu, 2015) are indicators of Economic Growth influencing the growth rate of an economy (Boldeanu, Constantinescu, 2015), human and physical accumulation (Moral Benito E, 2012).Another indicators such as capital formation, government expenditure (Barro, 1991)and it is strongly related to the public investment share unemployment rates have different impacts on the economy. Some indicators have major influence than others (Boldeanu, Constantinescu, 2015). Consumer Expense Index, Total Factor of Productivity are both considered one of the critical economy indicators. Total Factor of Productivity (TFP) measures the ability of a country to create and sustain a decent employment opportunity. Since 1957 TFP has reminded a major source of growth (Helpman, 2004). Countries with a high human capital rate tend to grow faster due to the rapid rate of introduction of new goods (Barro, 1991).

METHODS

This section is prepared as follows. Initially we discuss the variable selection and data. Then, the statistical and visualization methods employed was introduced. The data was collected from trusted sources and limit the date range to determine the economic growth indicator that reflected a slowdown in the economy during and after the Pandemic. Useful, and important indicators are present in this paper such as GDP, CIP, TFP, and Inflation. However, were limited in the measurement of the data extracted from the websites. The main source of data were the World Bank Data, the OECD website, and the Conference Board databases. Based on the literature survey and international reports, a set of economic indicators such as TFP, Capital Formation, Human Capital, and R&D were decided to investigate the effects of COVID-19 during the years of 2016-2021. Scope of data analysis were kept to OECD countries due to availability and reliability of the data sources. Table 1 depicts the overview of data, where variables, availability period and sources are provided. Furthermore, descriptive statistics of the data is provided in Table 2.

Table 1. Data Overview

Indicators	Unit of measurement	Availability Period	Source	Importance of Indicator	Reference
Final Consumer Expenses	US dollar	2016-2020	The World Bank Database	Primary	(Boldeanu, Constantinescu, 2015)
Natural Resources	% of GDP	2016-2020	The World Bank Database	Secondary	(Boldeanu, Constantinescu, 2015)
Technology and Research	% of GDP	2016-2020	OECD Database	Primary	(Contantinecu,2015)
Trade	US dollar	2016-2021	The World Bank Database		(Wolla, 2018)
Inflation	% of Inflation	2016-2021	The World Bank Database	Primary	(Callen)
Gross Capital Formation	US dollar	2016-2021	The World Bank Database	Secondary	(Boldeanu, Constantinescu, 2015)
Capital Share	GDP	2016-2021	The World Bank Database	Primary	(Hicks, 2020)
Total Factor of Productivity	Growth of the TFP	2016-2021	The Conference Board	Primary	(Romer, 1986)
Consumer Price Index	Annual Percentage	2016-2021	The Conference Board	Primary	(Masterson,2022)
Gross Domestic Product	US dollar	2016-2021	The World Bank Database	Primary	(Deninson, 1962)
General Government Final Consumption	US dollar	2016-2021	The World Bank Database	Secondary	(Boldeanu, Constantinescu, 2015)
Unemployment	% of labor force	2016-2021	The World Bank Database	Primary	(Jackson,2021)

Table 2. Descriptive Statistics

	Final Consumption Expenditure	**Technology and Research**	**Core Inflation Index**	**Total Factor Productivity**	**Consumer Price Index**	**Trade**	**General Government Consumption Expenditure**	**Unemployment**
Mean	1,294,475,239,835	1.67	2.06	-0.05	3.01	$1,472,835,873	$248,374,295,712	6.70
Standard Error	236,731,393,286	0.09	0.17	0.16	0.60	$7,874,598,595	$35,248,786,097	0.25
Median	254,865,066,472	1.37	1.71	0.21	1.85	$3,455,292,567	$80,545,527,906	5.69
Standard Deviation	3,347,887,470,239	1.16	2.50	2.30	8.58	$115,732,490,918	$518,049,239,937	3.71
Kurtosis	20	0.74	20.01	7.90	83.92	25	$20	3.75
Skewness	4.45	0.93	3.79	-1.84	8.90	-4.25	4.24	1.78
Range	20,902,843,381,159	5.43	20.84	19.42	89.17	$1,105,642,215,116	$3,348,982,765,314	21.53
Minimum	1,145,251,396	0	-1.25	-13.94	-1.93	$(845,050,000,000)	$4,748,234,686	2.01
Maximum	20,903,988,632,556	5.43	19.60	5.48	87.24	$260,592,215,116	$3,353,731,000,000	23.54
Missing data %	9.91%	27.48%	2.70%	8.56%	8.56%	2.70%	2.70%	0%

In terms of the methods of analysis, statistical t-test were utilized to compare pre and post pandemic periods. Pre-pandemic period was taken as the time between 2016 and 2019, then post pandemic period included the 2020 and 2021 years. The most up to date data that is available of all of the indicators was in 2021 therefore, 2016-2021 period was investigated.

When analyzing the data, even though OECD economies are close in various economic metrics, several outliers were identified by using Z score normalization. The data points that were more than 2 standard deviation away from the mean were excluded in the analysis to have a better understanding about the overall trends in pre vs post pandemic periods. In addition to the statistical t tests, box plots were used as the main statistical data visualization method to compare and contrast the pre vs post pandemic periods on the selected economic indicators. Therefore, those countries with outlier data points were excluded from the boxplot models. When creating the boxplots, we notice the upper and lower whiskers suggesting that these outliers were statistically significantly different from the other countries. Next section provides the results of our investigation.

RESULTS

The COVID-19 pandemic had profound effects one economies around the world. Governments and industries faced unprecedented challenges as they were facing the virus, implementing stricter measures and facing the socio-economic consequences. This paper examines the economic trend in OECD countries from 2016 to 2021. The pandemic effects on the economy has been extensive across the various sectors of the economy. One of the most striking effects was in the Unemployment rates, Inflation, Total Factor of Productivity Technology and Research Investment index, Government Final Consumption Spending and Consumer Price Index. Initially the results of statistical analysis are provided in Table 3.

Table 3. Results of Statistical T-test: Two-Sample Assuming Equal Variances

Variable	P(T<=t) one-tail	P(T<=t) two tail	Change is Significant (95% CL?
Final Consumption	0.230614879	0.461229758	No
Natural Resources	0.107324052	0.214648105	No
Technology and Research	0.20264004	0.405280081	No
Inflation	0.021175962	0.042351923	Yes
Gross Capital Formation	0.25464773	0.509295461	No
Capital Share	0.032437843	0.064875687	Yes
Total Factor of Productivity	0.001145087	0.002290173	Yes
Consumer Price Index	0.187984998	0.375969996	No
Gross Domestic Product	0.295858065	0.59171613	No
General Government Consumption Expenditure	0.15967671	0.319353421	No

Unemployment	0.060940366	0.121880732	No
Trade	0.283576847	0.567153693	No

Results indicate that inflation and capital share, and total factor productivity worsened significantly in the post pandemic era and the changes were statistically significant based on an alpha value of 5%. Rest of the indicators did not result in a significant change, but boxplots indicate that majority of them also either stay stagnant or worsened.

By understanding the economic trends and difficulties faced by OECD countries during COVID19, policymakers would be able to make informed decisions to facilitate a resilient a sustainable recovery. Following sub sections provide the results of boxplot visualizations.

Trade

The trends of trade is depicted in Figure 1. We noticed a decline in trade in 2016, while in 2017 increased by a small percentage. In 2017 The World Tarde Statistical Review recognized a higher Merchandise Trade, which was up sharply from 2016 since the trade volume increased by 1.8% (Trade Statistics - World Trade Statistical Review, 2021). In 2018, there is an increase in trade because of export orders that are strong predictors of merchandise trade volume. Consequently, a small constant decrease in 2019-2020 where travel, and trade restrictions were implemented. The WTO forecast about global trade predicted a faster recovery in 2020. In 2020, we can notice a slowdown in the economy, and slow levels in the OECD trade. Driving the global economy to a slow deterioration, holding high tariffs. However, global trade in 2020 was not severe as was estimated by the World Trade Organization.

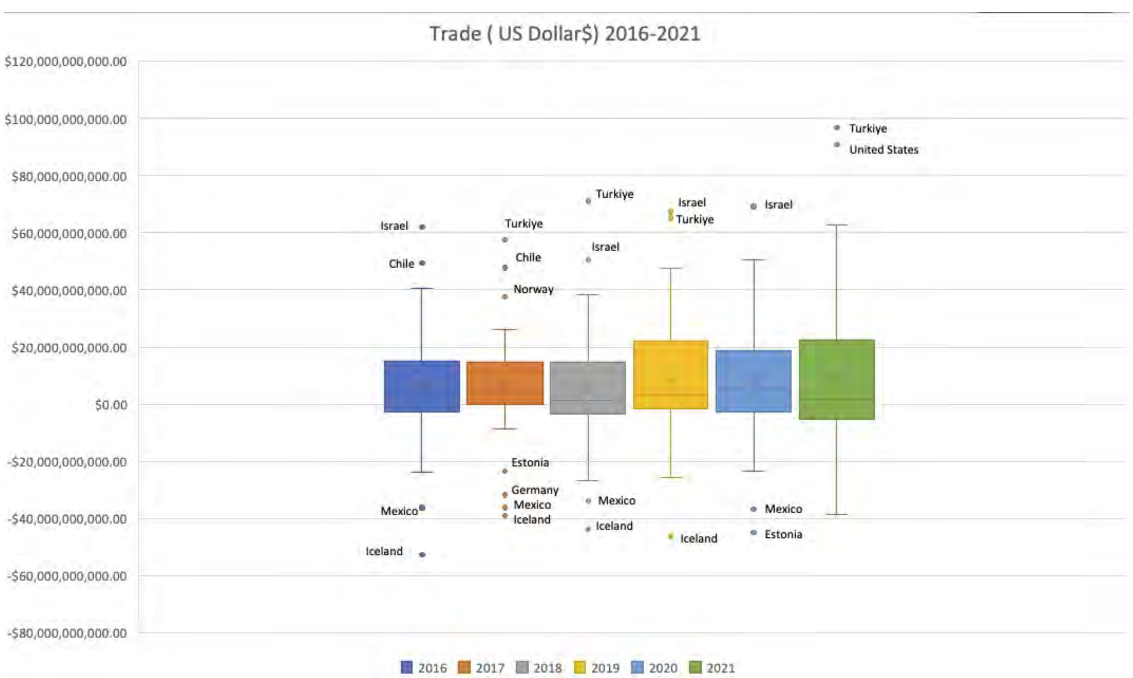


Fig. 1 OECD Trade, data are in current U.S dollars from 2016-2021.(Source: International Monetary Fund, Balance of Payments Statistics Yearbook and data files.)

In 2017, the United Kingdom trade percentage was 62.79 and during the next two following years increased by 2.11878%; however, 2021 showed a decline of 11.6323%. The United Kingdom stands out by its export of cars, gas turbines, crude oil, and medicaments, all of these items were highly demanded during this year. According to the WTO Trade outlook, world trade fell by 13 and 32% because of the pandemic, the WTO believed it was most likely to exceed the 2008-2009 recession number, specifically because of the short supply of automobiles. Consequently, trade services such as transport and travel were highly impacted by the government and health restrictions imposed, such as the GDP falling 4% on average.

Table 1 The impact of COVID-19 on global trade and % change between the years 2016-2021

Country/ Year	2016	2017	2018	2019	2020	2021
Luxembourg	348.43	353.79 (1.53%)	362.42 (2.43%)	377.84 (4.25%)	365.21 (-3.34%)	388.12 (6.27%)
Ireland	227.73	220.00 (-3.39%)	217.07 (-1.33%)	252.24 (16.20%)	247.76 (-1.77%)	229.44 (-7.39%)
Slovak	184.03	188.05 (2.18%)	189.80 (0.92%)	183.47 (-3.33%)	168.51 (-8.15%)	187.82 (11.46%)
Hungary	164.40	165.22 (0.50%)	163.26 (1.19%)	160.78 (-1.51%)	155.52 (-3.26%)	162.79 (4.67%)
Slovenia	146.65	157.27 (7.23%)	161.14 (2.45%)	158.79 (-1.45%)	146.23 (-7.91%)	160.92 (10.04%)
Italy	55.36	58.60 (5.84%)	60.30 (2.89%)	59.87 (-0.70%)	55.28 (-7.66%)	62.98 (13.92%)
Australia	40.82	41.95 (2.76%)	43.38 (3.42%)	45.82 (5.61%)	44.23 (-3.46%)	39.87 (-9.86%)
Colombia	36.20	35.28 (-2.54%)	36.53 (3.54%)	37.55 (2.80%)	33.90 (-9.72%)	40.85 (20.85%)
Japan	31.31	34.42 (9.94%)	36.63 (6.41%)	35.17 **(-3.96%)**	31.35 **(-10.86%)**	37.37 (19.20%)
United States	26.58	27.27 (2.60%)	27.61 (1.21%)	26.45 (-4.19%)	23.28 **(-11.59%)**	25.48 (8.97%)

Source: International Monetary Fund, Balance of Payments Statistics Yearbook and data files.

As shown in the Table 1, Luxembourg is the first in the category by being one of the top exports of Iron Blocks ($1.37B), Cars ($696 M), Rubber Tires ($677M), and Iron Sheet Piling ($469M), mostly to Germany (OEC Luxembourg, 2021). Australia is known for being the fifth larger exporter of pearls, zircon, and silver. The Australian Bureau of Statistics stated that in February 2020 the god and services credits, non-rural goods, and non-monetary gold fell by 7 %. However, in 2021 the Australian Parliament declared that imports were heavily impacted by the Pandemic and decreased by 9.864% from 2020. Trade represents a large proportion of the Australian GDP since there is no local or short-distance trade due to Australia being an island nation. During 2020, all of these countries from the sample experienced a decline in their exports, which showed a negative percentage change between 2019 and 2020. In 2021 there was a positive percentage increase in trade suggesting that the economy was improving.

Inflation

Figure 2 depicts the inflation rate trend. In this model the core inflation index is measured as the percentage of inflation. Between 2016-2017 and 2018, there is a change of 0.25 inflation rate. In 2019, there are lower levels of inflation not overpassed 1.4 of the core inflation indexes, because of the COVID-19 emergency many governments assessed short- and long-term costs and benefits for the population, to address job dislocations as a form of relieving the economic weight. The recession brought high levels of unemployment and affected lower-wage and less-educated workers. However, in 2021 their economy is growing, showing high levels of inflation because of the overconsumption, and the economic monetary policies that were implemented. It is important to notice that there is a normal level of inflation since there is a global average inflation in 2020 increase of 2.8. A clear example of how prices went up in less than a year, is the crude oil barrel price. Oil prices went from $20 per barrel in April 2020 and went up to $40 per barrel by the end of the same year. In the following year, in 2021 the price of crude oil went from $70 per barrel to $80 per barrel. Additionally, the Inflation index was increasing because of the energy and oil prices, higher demands, but supply shortage.

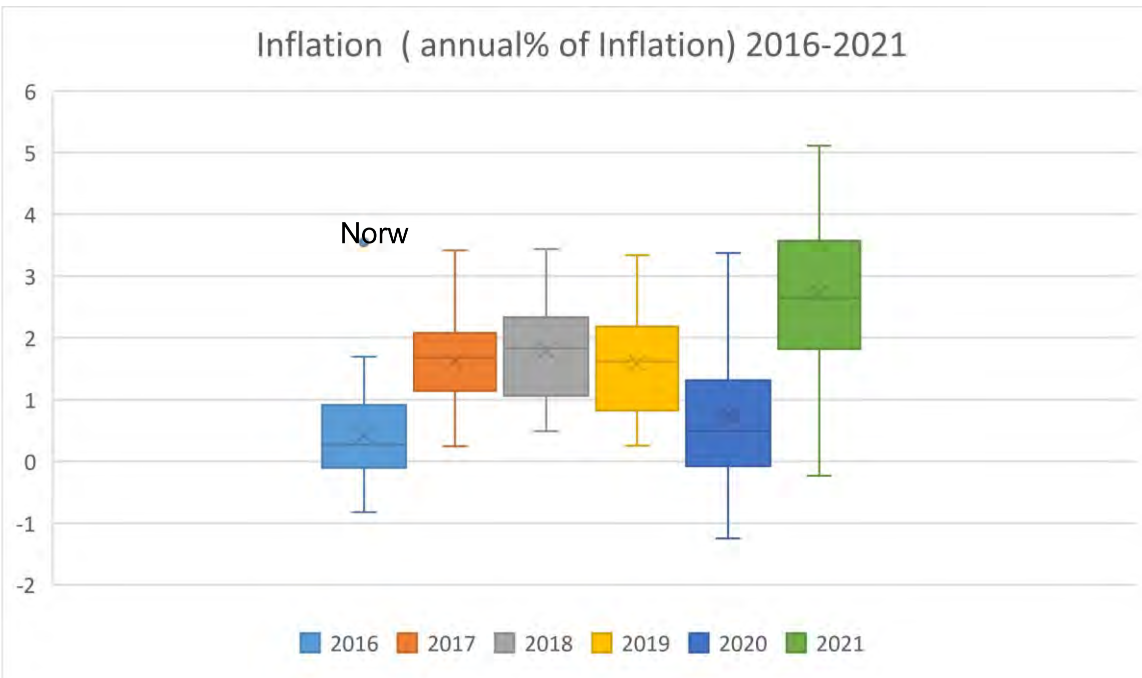


Fig. 2 Inflation reflected by the annual percentage change in the cost of goods and services from 2016-2021. **Source:** International Monetary Fund, International Financial Statistics, and data files

Unemployment

The unemployment rate (Fig. 3) showed a decline since 2016, as per the lowest unemployment too. It is assumed to be the percentage of the natural rate of unemployment in the long term. However, the rate of unemployment went as high as 6.9 globally. Then, in 2020 and 2021 the unemployment rate has been declining, returning to its natural rate of unemployment. During 2019-2020 governments responded with many policy initiatives such as an unemployment insurance system that provided short-term unemployment insurance to assist workers' income.

Also, for 2019-2020 years many workers were forced to leave the workforce because of the schools and childcare closing, and there was an increase in retirement due to health concerns and unemployment insurance programs these programs assisted workers by comprising the increased benefits for existing programs that were designed to support works that lost work hours to maintain pre-pandemic employment income levels.

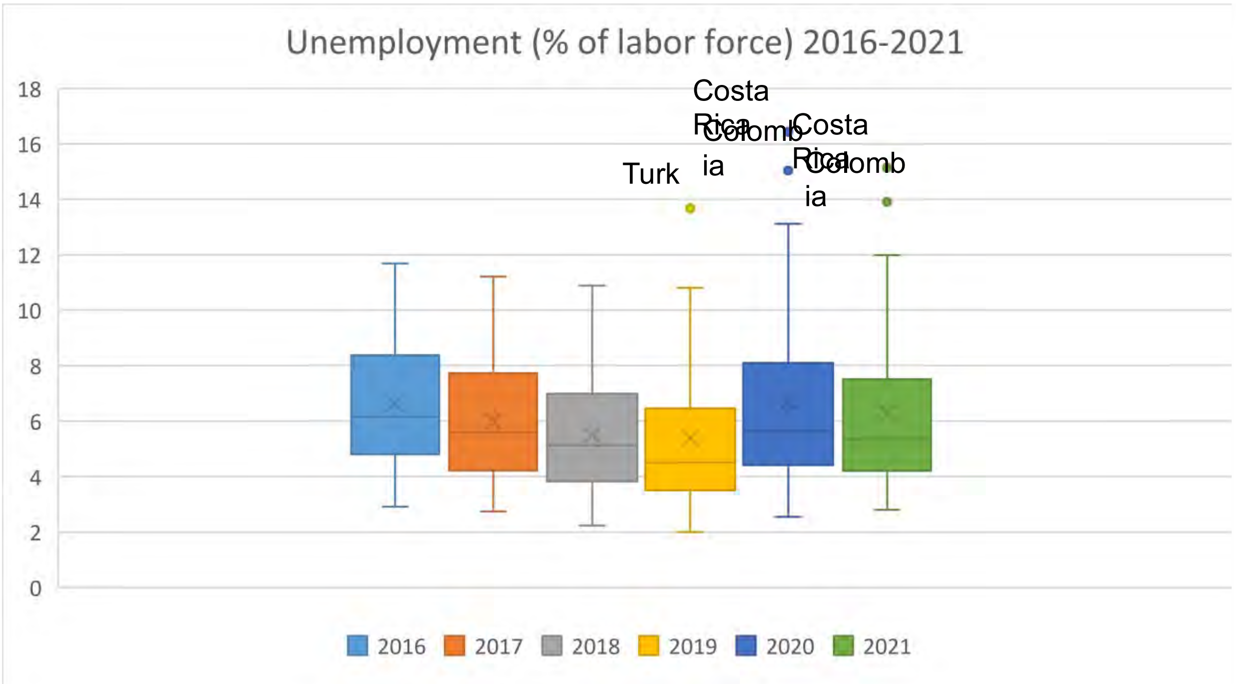


Fig. 3 Unemployment, total measure in % of labor force from 2016-2021(modeled ILO estimate) **Source:** International Labour Organization. "ILO Modelled Estimates and Projections database (ILOEST)" ILOSTAT.

During 2019 the unemployment rate increased as per the concerns of mortgages and rent prices could increase, financial institutions were needed to provide a mechanism or financial assistance. The International Labor Organization estimated that unemployment around the world was equivalent to 0.9% of the work hours that were lost in 2020. Additionally, the ILO estimated that the highest working hours lost were in Europe due to the stricter lockdowns and work restrictions. The lost work hours were 3.3% higher than in America. Moving to 2021, we faced high levels of inflation as per higher levels of unemployment, it was a matter of time until the 2021 unemployment rate gets back to its natural unemployment rate. In 2021 The ILO estimated that there was a loss of 5.3 % in the global worker income. Likewise, the OECD also announce that the recession of 2020 cost 22 million jobs in the countries that are part of the OECD. It was also determined that people that lost their jobs or were laid off by their employers socially were divided by their skill level, education, or gender basis, as a result, the OECD worked toward closing the inequality breach in many OECD countries.

Final Consumption Expenditure During 2018 and 2019, there was a slight increase in consumption expenditure. In 2020 there was a higher level of expenditure reflecting the income increase and support from the government policies. In 2021, in the US the Federal Reserve communicated that there was a slowdown in economic activity because of the low levels of consumer dining and personal consumption activities this was a rising because of the virus spread concerns. Also, many wholesale and resellers experienced a low level of production because of labor and supply shortages, all of these supply shortages lead to low inventory levels that made it harder for people to consume at average prices, leading to consumers spending money on overpriced items.

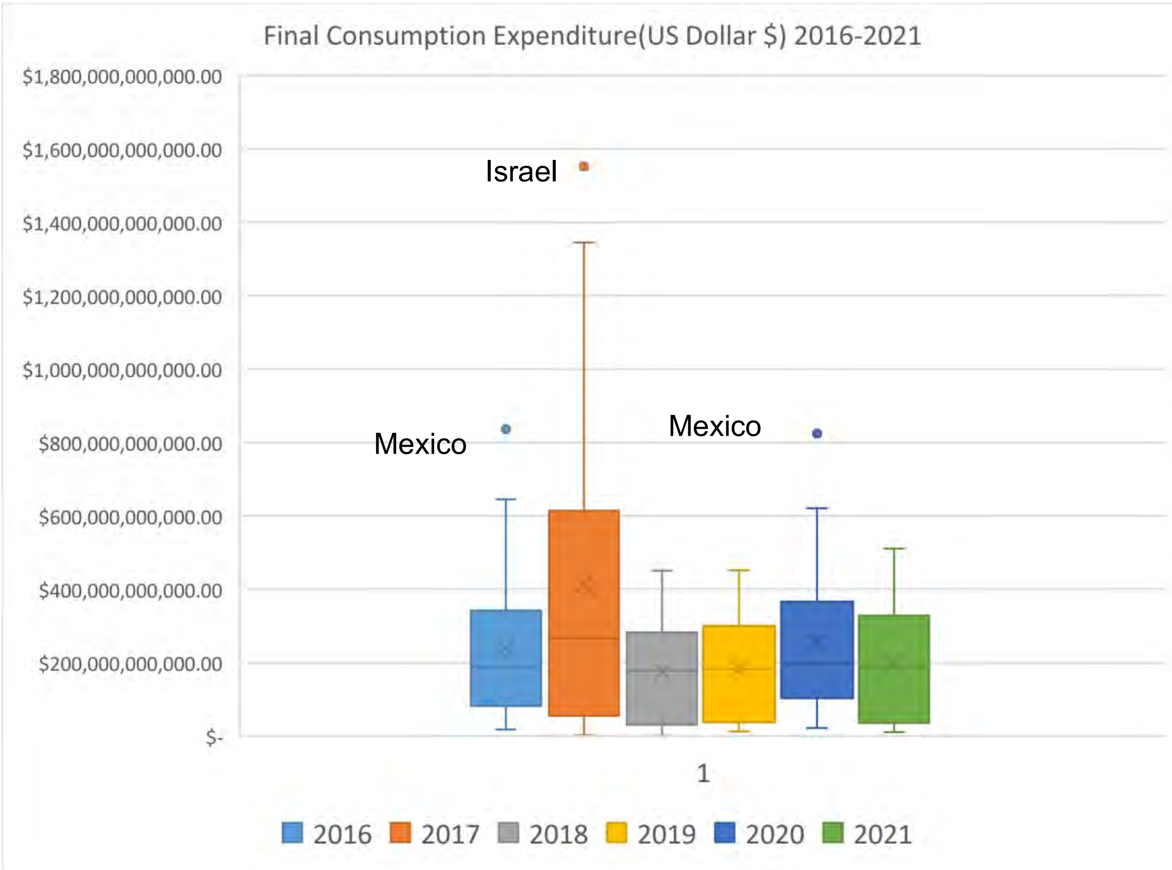


Fig. 4 Final Consumption Expenditure measure in current US dollar from 2016-2021. **Source:** World Bank national accounts data, and OECD National Accounts.

During 2019 and 2020 many consumers from Europe and the U.S. adjusted to the pandemic strict restriction and were confident about the Government unemployment benefits, personal savings, and loans from the Bank to maintain their consumption of necessities. Because of the house lockdowns that the government implemented, many household expenses were increasing.

In 2021, the consumption expenditure slightly decreased, the International Monetary Fund forecasted a decline in it because consumers cut their spending and increased their savings as part of concerns of a recession. The IMF also forecast that countries from all over the world were experiencing side economic effects due to the Pandemic such as the health crisis, domestic economic crisis, external demand, capital outflow, and the collapse of price levels. All of these effects were measured by different criteria such as the length of required lockdowns in different countries, impact on firm closures and unemployed workers leaving the workforce, and global supply chain that affected companies' productivity as per supply disruptions. On the other side, we can notice the low consumption level in 2021, the U.S. Bureau of Economic Analysis released an analysis of the deceleration in the growth rate that showed lower levels of consumption in goods and services. As part of the result, the U.S. GDP fell 5 % at its annual rate in the first quarter of the year. Compared to 2020, which did not fall as much as in 2021, 2020 the annual growth rate fell 3.4%, as per 2019.

Germany on the other side on 2019-2020 fell -3.7% in economic growth. An OECD Economic survey on Germany stated that during the first half of 2020, the economy contracted sharply, as there was a failure in private consumption and exports (OECD Economic Survey: Germany, 2020). In June 2020, a comprehensive economic stimulus pack was presented with important investments, the recovery package was intended to boost consumption, including a reduction in the value-added tax rate that went from 19% to 16 %, the from 7% to 5 %. The German Recovery and Resilience Plan contributes to overcoming the Covid-19 effects and protecting of Germany and Europe (Federal Ministry of Finance, 2020).

Technology and Research

In Figure 5, we can notice 2016 to 2018 a constant decrease in technology, as the mean was 1.26. In 2019 there was a slight increase in technology by 0.2 %. The Main Science and Technology Indicator showed the statistical publication on the level of the structure of the efforts of the OECD countries on financial and human resources that devoted to research and experimental development. From 2019 and 2020 there was an increase in Research, Human, Financial, and Development resources % growth. Besides Human and Financial Resources, patent data and international trade in R&D-intensive industries taken into account R&D intensive industries pharmaceuticals, computer electronics, scientific research, and development. During 2019-2020 there was a % increase meaning that during the Pandemic many companies spent most of their time and funds expanding their Research & Development, mostly digital and pharmaceutical companies. On the other side, the Pandemic opened the door for telework, elearning, and e-commerce. It is important to notice that even though there was an economic slowdown, these activities kept the economy going. Between March and September 2020 many organizations such as the Government and Philanthropic organizations worldwide, invested around USD 5 billion in R&D COVID-19 (OECD calculations, 2020). In the Asia Continent, there was a USD 300 Million investment, USD 800 million in Europe, and around USD 3.5 billion in North America (OECD, 2020).

Governments from Europe told their employers to send their workers to stay at home during the Pandemic. The Europe and Investment Bank provided 15 million euros to IRBM and Italian Specializing in research and development for coronavirus and infectious diseases research. They focused on vaccine production and quality controls (Squintani, 2022). As well as Japan 2021 created a new fund to increase the financing for Scientific research, the pandemic showed there was a lack of research funding interest from the government, the new fund is under the Japan Science and Technology Agency and would increase the R&D infrastructure (Squintani, 2022). Indeed, COVID-19 changed the way people work, study, and interact. Because of the Pandemic, many companies and learning institutions have accelerated digitalization.

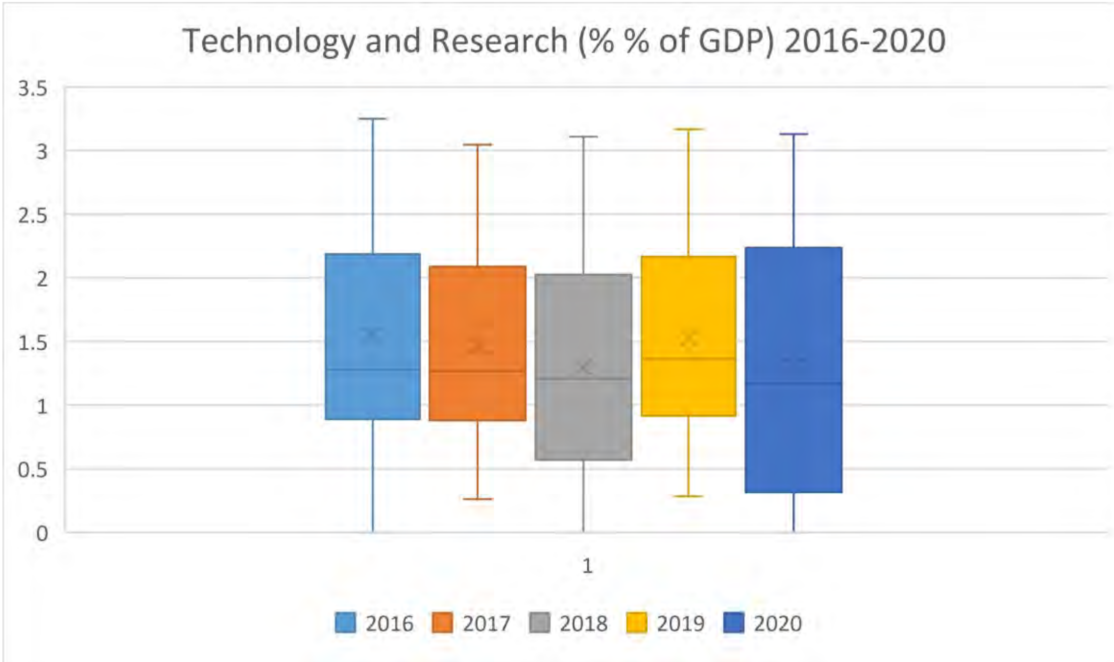


Fig. 5 Technology and Research measured in GDP % from 2016-2020 **Source:** Data derivate from national R&D surveys and budget supplied by the national statistical agencies to the Secretariat vi an OECD/Eurostat collection

Total Factor Productivity

Productivity is a measure of how efficiency is in an economy, it takes into account as an input labor, capital, land, energy, and everything that is used to produce goods and services (Kapsos, 2021) . In figure 6, we can notice the sharp decline in the global Total Factor of Productivity in 2020. From 2018 there is a slight decline, while from 2019 to 2020 went to -2.5. On the positive side, in 2021 an increase setting the TFP at 2.5.

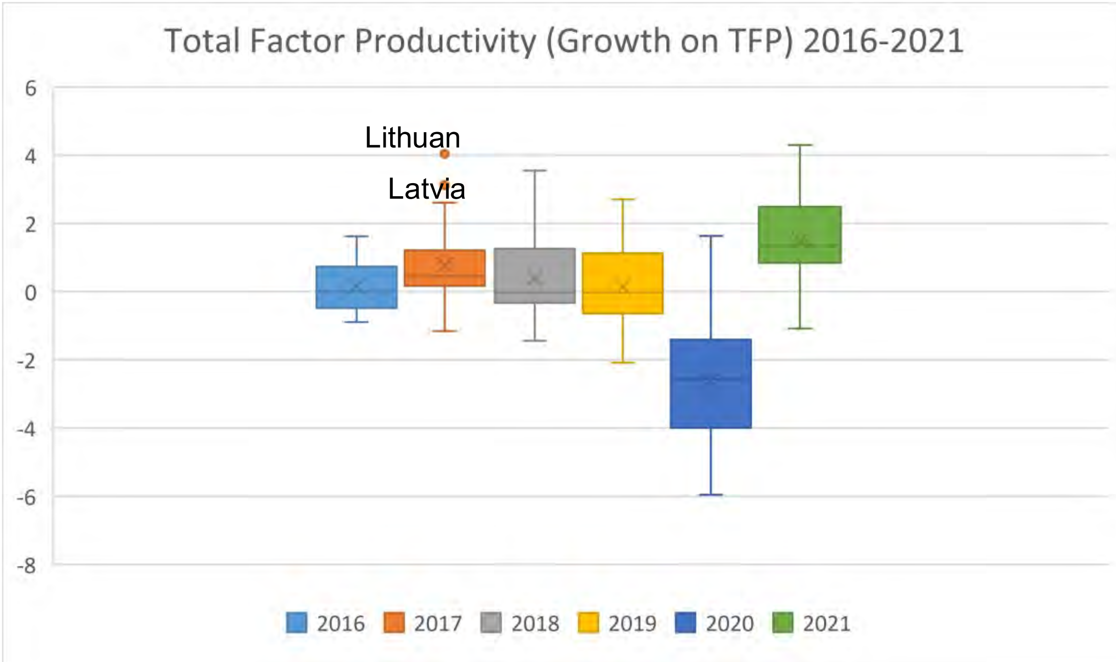


Fig. 6 Total Factor Productivity measure in growth on TFP from 2016-2021 **Source:** The Conference Board Website Database

The economic effects were reflected in the production lines that suffered from shutdowns and border closures, many factories were closed as social distancing was strictly implemented. As factories were shut down, many companies were evaluating if the risk exposure was worth it for the supply chains for basic necessity items that were important for national security. As a result, all the global supply chain changes and relocations affected productivity as the companies tried to keep their business safe. We notice how the OECD productivity fell in 2020 but recovered through 2021. Therefore during 2020, there is a sharp decline, the Conference Board released an analysis were showed that TFP per worker fell by 0.9. The acceleration of digital work, investments, new infrastructure, and human capital helped boost the worker's productivity. Based on the Conference Board record from 2020(The Conference Board, 2021)

The United States in 2020 because of the COVID-19 shutdowns fell -5.8 % in productivity, while in 2021 increased 4.7, even if work hours increased the unemployment rate still is one of the highest that the US has experienced since 2009. Between France, Italy, and Germany, Spain dropped its productivity by -0.6%, and Germany fell slightly. The United Kingdom managed to remain without drastic changes in its productivity due to its strong employment performance; however, its hours worked decreased in 2021. Unemployed affected labor productivity as employment rates decreases, and the employment age was the highest historically in the OECD countries. Based on the OECD Labor Market Statistic, the pandemic decreases the employment rate by 2.7% between 2019 and 2020 (OECD National Accounts Statics database, 2021)

Consumer Price Index

We identified a sharp decline in the consumer prices during 2020 (See Fig. 7). We can notice the trend and constant decrease in price from 2017 to 2020. However, we notice a rapid decrease in 2020 with a mean of 1.6 while the constant mean from 2017-2019 was 1.9. We can see a sharp decline in consumer prices during 2020. We can notice the trend and constant decrease in price from 2017 to 2020. However, there is a rapid decrease in 2020 with a mean of 1.6 while the constant mean from 2017-2019 was 1.9. The 2021 Europe economic forecast of the European Commission stated that Europe's GDP would rise by 4.8% in 2021; however, they still anticipated a slowdown in the economy as a result of the business lockdowns. The forecast also anticipated a faster recovery in travel and tourism, which would reflect higher consumer spending levels. On the seasonal Consumer Price Index publication from the U.S Bureau of Labor Statistics in 2020 the CIP experience a lot of changes due to the pandemic, as gasoline prices decrease by 53.0 cents in 2020. Food prices were highly affected by the Pandemic because of the lockdowns and strict quarantines that force people to order delivery services and normally go to the supermarket and purchase their food. It affected the purchase and production of it (Congressional Research Service, 2021).

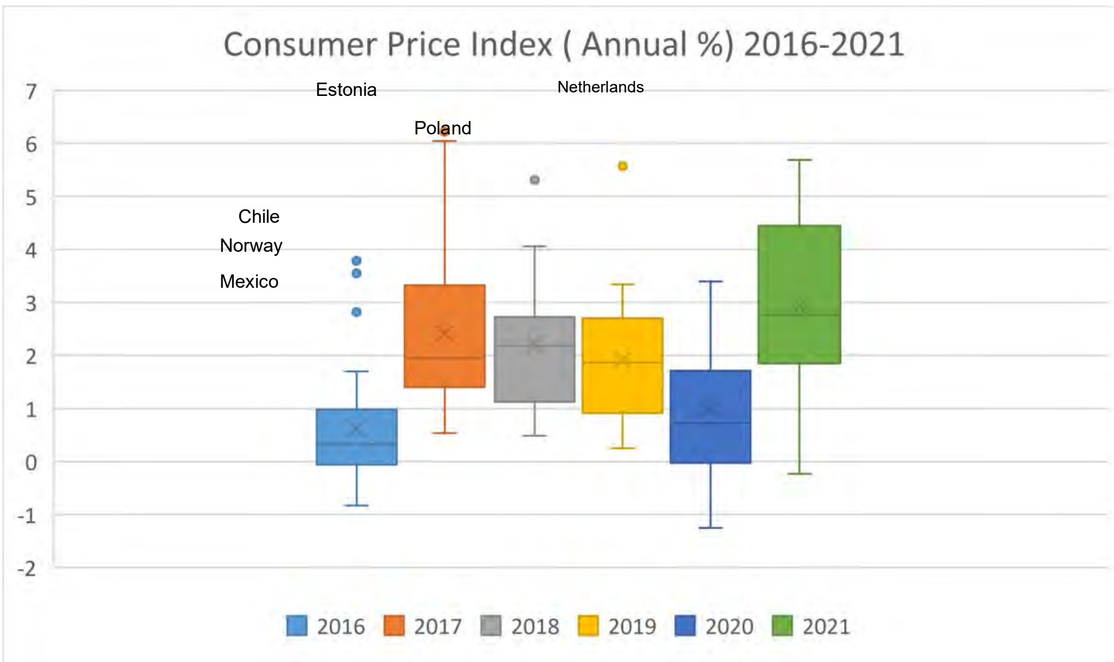


Fig. 7 Consumer price index measure in annual % from 2016-2021 **Sources:** The Conference Board Website Database.

Food prices became more volatile, during the year prices were rising 0.5, 2.7, 0.8, and 0.5 %. As the price change increased the final demand and the Costumer Price Index for food homes were controlled by high volatility, especially for meat, dairy, and eggs (U.S. Bureau of Labor Statistics, 2022).The Automobiles market was affected and surprisingly the used vehicles. The production of vehicles decreased because of the factory shutdown. Consequently, dealers were having short supply issues making the ones that were in inventory more price valuables, making car prices higher and reducing sales. Figure7 shows the effect of food and cars price increase in 2021 bar, with the mean going from 0.7 in 2020 to 2.7 in 2021. We also, notice one of the lowest pitfalls of gasoline, gasoline prices are usually volatile, but covid-19 was a cause of extreme price volatility in gasoline. The CIP fell sharply and went up in 2021 making it almost inaccessible for low-income families (Blake,2022)

General Government Consumption Expenditure

In Figure 8, the General Government's final consumption is measured in US Dollars $. The model shows a constant increase in the Government Spending. Countries like Japan increase their National Budget to help the economy to keep working, Japan implemented a monetary policy where the Bank of Japan kept a low-interest rate of -0.1%. Also, the Japan COVID-19 lending facility helped banks to provide zero-interest rate loans from small to big businesses for them to help them to ensure salaries to their employees and pay the bills. The United States is known for providing $4.6 Trillion (GAO, 2023) to help the nation to overcome and deal with the COVID-19 economic effects. Also, the United Kingdom spend £310 billion to £410 billion. This is the equivalent of about £4,600 to £6,100 per person in the UK (Brien, 2023). The money was used on public service, support for small business, and individuals.

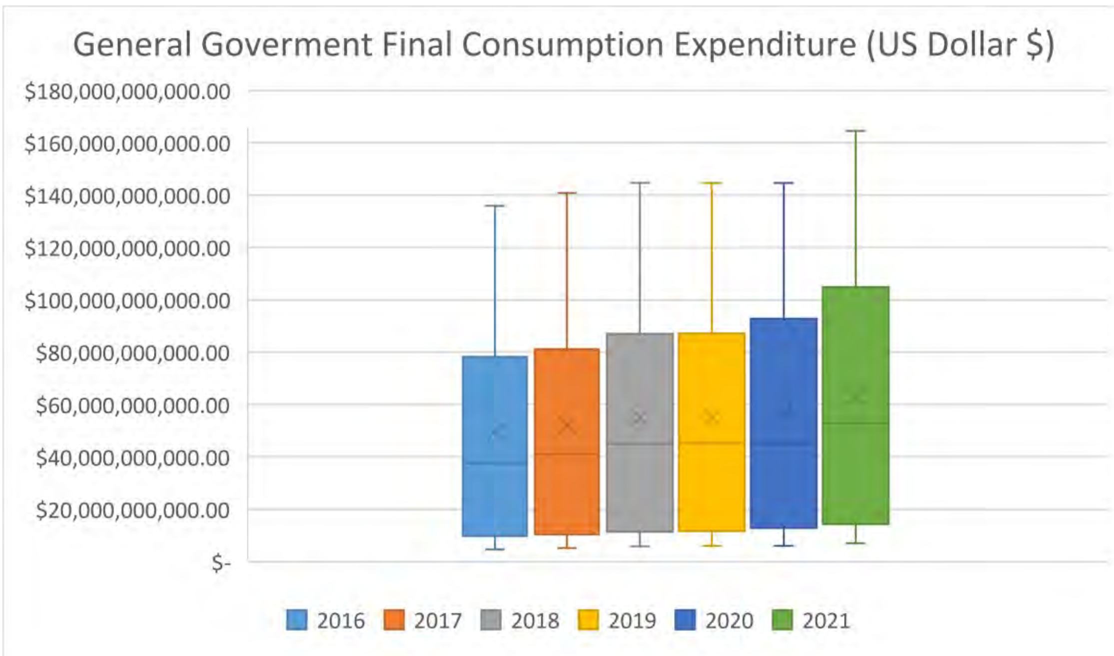


Fig. 8 General Government Final Consumption Expenditure measure in current US dollar from 2016- 2021. **Source:** World Bank National Accounts data, and OECD National Accounts data files.

CONCLUSIONS

In conclusion, the COVID19 pandemic had an outstanding impact on the economy, that led to high unemployment rate, supply chain shocks, spikes in consumer prices, lower productivity and significant declines in the government and household consumption. On the other side, there were also opportunities of innovation such as an increase in research and development funding, and the acceleration of digitalization (work from home, e-learning). Therefore, by implementing the recommendations presented in this paper, countries would be able to move towards a resilient and sustainable recovery period. Governments have a ton of work to restabilize the world economy, foster employments opportunities and sustainable creation of an inclusive growth. This study had certain limitations. The findings presented in the graphs relied on data extracted from OECD and World Bank databases, which introduced inherent variations that made the analysis and interpretation time-consuming. Furthermore, the availability of data and information related to global Consumer Price Index and Total Factor of Productivity indicators was restricted, primarily due to their predominant usage within the United States.

The research's findings showed that, there have been numerous serious financial challenges. Among these difficulties were the consistently high unemployed rates that have affected economies all across the world. Disruptions in the supply chain have also become an important concern, pushing up consumer costs and causing inflationary pressures. Lower production levels and significant drops in government and household consumption have followed as a result. Therefore, governments face a challenging responsibility in increasing employment opportunities and restoring stability to the global economy while simultaneously promoting sustainable and inclusive growth. The complicated terrain of the post-financial crisis recovery requires a coordinated effort to address these multifaceted concerns.

Relationship between OECD Countries and Emerging economies

OECD countries support and promote inclusive and sustainable employment creation, which are crucial goals for developed and emerging economies. Emerging economies are the ones that usually have lower income levels in comparison to developed economies, where sustainable and inclusive job creation is encouraged as it will boost productivity and consumption creating a stronger economy. Additionally, emerging economies were impacted by supply chain shocks because EE's heavily relies on exports as a source of income, usually resulting in economic contraction. On the contrary, EE's most of the time are suppliers of input materials, and when developed countries experience disruptions in supply chains the number of orders is reduced impacting manufacturing and economic activity in both markets. Lastly, when consumption decreases in OECD countries, the demand for commodity prices decreases, in consequence as emerging economies are major exports of commodities such as metals and minerals.

REFERENCES

Amitrajet A. Batabyal. (2022). *How total factor productivity explains how nations become wealthy | World Economic Forum*. Why Some Nations Are Richer than Others. https://www.weforum.org/agenda/2022/07/wealth-nations-some-richer-than-others/

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, *106*(2), 407–443. https://doi.org/10.2307/2937943

Boldeanu, F. T., & Constantinescu, L. (2015). The main determinants affecting economic growth. *Bulletin of the Transilvania University of Braşov Series V: Economic Sciences •*, *8*(57).

Coryanne Hicks. (2021). What Are the Four Factors of Production? *What Are the Four Factors of Production*. https://money.usnews.com/investing/investing-101/articles/what-are-the-fourhttps://money.usnews.com/investing/investing-101/articles/what-are-the-four-factors-of-productionfactors-of-production

Federal Ministry of Finace. (2020). German Recovery and Resilience Plan – Draft. In *German Recovery and Resilience Plan*.

Frenkel, J. A., & Razin, A. (1986). Fiscal Policies in the World Economy. *Https://Doi.Org/10.1086/261390*, *94*(3, Part 1), 564–594. https://doi.org/10.1086/261390

Gokhan, E. (2023). Syllabus MGMT Independent Study.Lindenwood University, Dept. of Management, St. Chales, MO

Islam, M. S., & Alhamad, I. A. (2022). Impact of financial development and institutional quality on remittance-growth nexus: evidence from the topmost remittance-earning economies. *Heliyon*, *8*(12), e11860. https://doi.org/10.1016/J.HELIYON.2022.E11860

Jackson, J. (2021). *Global Economic Effects of Covid-19*. https://apps.dtic.mil/sti/pdfs/AD1152929.pdf

Kapsos, S. (2021). *Why would labour productivity surge during a pandemic? - ILOSTAT*. Why Would Labour Productivity Surge during a Pandemic? https://ilostat.ilo.org/why-would-labourhttps://ilostat.ilo.org/why-would-labour-productivity-surge-during-a-pandemic/productivity-surge-during-a-pandemic/

Kunal Sen. (2021). *UNU-WIDER : Blog : Why should I care about economic growth?* Why Should I Care about Economic Growth? https://www.wider.unu.edu/publication/why-should-i-care-abouthttps://www.wider.unu.edu/publication/why-should-i-care-about-economic-growtheconomic-growth#:~:text=Economic%20growth%20increases%20state%20capacity,protection%20and%20basic%20public%20services.

Mihi Ramirez, A., Sobierajc, J., & Garcia- Rodriguez, Y. (2019). *EconStor: Interaction of emigration and immigration with foreign direct investment, international trade and remittances*. Interaction of Emigration and Immigration with Foreign Direct Investment, International Trade and Remittances. https://www.econstor.eu/handle/10419/209123

Moral-Benito, E. (2012, May). *Determinants of economic growth: a bayesian panel data approach*. https://www.jstor.org/stable/23262088

National Accounts at a Glance. (2021). *Budgetary and fiscal policies //General government - General government spending - OECD Data*. General Government Spending. https://data.oecd.org/gga/general-government-spending.html

O'brien Gregory. (2020). *Australian trade in figures – Parliament of Australia*. Australian Trade in Figures.
https://www.aph.gov.au/About_Parliament/Parliamentary_departments/Parliamentary_Library/pubs/BriefingBook47p/AustralianTradeFigures

OECD. (2018). World Trade Statistical Review 2018 Latest trends in world trade. *World Trade Statistical Review Review 2018*.

OECD. (2020). *Business research and innovation have been affected unevenly by the crisis - OECD*. https://www.oecd.org/sti/science-technology-innovation-outlook/crisis-andhttps://www.oecd.org/sti/science-technology-innovation-outlook/crisis-and-opportunity/businessresearchandinnovationhavebeenaffectedunevenlybythecrisis.htmlopportunity/businessresearchandinnovationhavebeenaffectedunevenlybythecrisis.html

OECD Economic Outlook. (2023). *GDP and spending - Real GDP forecast - OECD Data*. OECD Economic Outlook. https://data.oecd.org/gdp/real-gdp-forecast.html

OECD organization. (2020). *OECD Economic Surveys: Germany 2020*. https://doi.org/10.1787/888934200261

OECD Stat. (2022). *Economic Outlook No 112 - November 2022.* Economic Outlook No 112 . https://stats.oecd.org/Index.aspx?DataSetCode=EO#

OECD World. (2022). *Luxembourg (LUX) Exports, Imports, and Trade Partners | OEC - The Observatory of Economic Complexity*. https://oec.world/en/profile/country/lux?latestTrendsScaleSelectorNonSubnat=Trade%20Value%20Growth%20Value

Peterson Institute for International Economics. (2012). Devaluing to prosperity: Misaligned currencies and their growth consequences. www.piie.com

Project invested. (2019, October). *How capital formation fuels business expansion, job growth and investment | Project Invested*. How Capital Formation Fuels Business Expansion, Job Growth and Investment. http://www.projectinvested.com/markets-in-action/how-capital-formation-fuelshttp://www.projectinvested.com/markets-in-action/how-capital-formation-fuels-business-expansion-job-growth-and-investment/business-expansion-job-growth-andhttp://www.projectinvested.com/markets-in-action/how-capital-formation-fuels-business-expansion-job-growth-and-investment/investment/#:~:text=%E2%80%9CCapital%20formation%2C%E2%80%9D%20in%20its,capital %20will%20take%20different%20forms.

Ruffin, R. J., & Gould, D. M. (2015). *What Determines Economic Growth? - Economic Review, Second Quarter 1993 - Dallas Fed*. What Determines Economic Growth? https://www.researchgate.net/publication/5029896

Squintani Lorenzo. (2022, April 21). *Italy: EIB provides €15 million to fund coronavirus research at IRBM*. Italy: EIB Provides 15 Million to Fund Coronavirus Research at IRBM. https://www.eib.org/en/press/all/2022-201-italy-eib-provides-eur15-million-to-fund-coronavirushttps://www.eib.org/en/press/all/2022-201-italy-eib-provides-eur15-million-to-fund-coronavirus-research-at-irbmresearch-at-irbm

The Conference Board. (2021, April 21). *Global Productivity Growth Remains Weak, Extending Slowing Trend*. Global Productivity Growth Rate Extending Slowing Trend. https://www.conference-board.org/press/global-productivity-2021#:~:text=Hourly%20output%20grew%20by%20%2B0.8,substantial%20%E2%88%920.6%20percent%20in%20Spain.

The Investopedia team. (2021, January 1). *What Is Economic Growth and How Is It Measured?* What Is Economic Growth and How Is It Measured? https://www.investopedia.com/terms/e/economicgrowth.asp

The Wold Bank Group. (2023). *Science & Technology | Data Expenditure*. Science and Tehnology. https://data.worldbank.org/topic/14

The World Bank. (2023a). *Personal remittances, received (% of GDP) - OECD members | Data*. Personal Remittances. https://data.worldbank.org/indicator/BX.TRF.PWKR.DT.GD.ZS?locations=OE

The World Bank. (2023b). *Total natural resources rents (% of GDP) | Data*. The World Bank Group. https://data.worldbank.org/indicator/NY.GDP.TOTL.RT.ZS

The World Bank Group. (2021). *(Aggregate Demand) Final consumption expenditure (current US$) | Data*. Final Consumption Expenditure (Current US$). https://data.worldbank.org/indicator/NE.CON.TOTL.CD

The World Bank Group. (2023). *International Trade/Imports of goods and services (% of GDP) | Data*. International Trade. https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS

Tofique, M. W., Löf, A., Millward, C., & Mou, J. (2020). IOP Conference Series: Earth and Environmental Science Experimental analysis of reed valve movement for different reed valve designs tested in an impact fatigue test system Research on the Impact of COVID19 on Global Economy. https://doi.org/10.1088/1755-1315/546/3/032043

YiLi Chien. (2015, June 1). *What Drives Long-Run Economic Growth?* What Drives Long-Run Economic Growth? https://www.stlouisfed.org/on-the-economy/2015/june/what-drives-long-runhttps://www.stlouisfed.org/on-the-economy/2015/june/what-drives-long-run-economic-growtheconomic-growth#:~:text=There%20are%20three%20main%20factors,Technological%20advancement

DECISION SCIENCES INSTITUTE

Is the Football Power Index a Good Bet?

Thomas R. Robbins
East Carolina University
robbinst@ecu.edu

ABSTRACT

Sports analytics have become ubiquitous in print, online and on television. Major media organizations with large, well-funded analytics departments have developed and promoted a type of predictive model known as a Football Power Index. In this paper we examine 3 models, the ESPN FPI, and two versions of the FiveThirtyEight.com Elo model. We discuss how they work, how accurate they are, and how they fare relative to sport betting markets. We also investigate what they tell us about semi-strong efficiency in sports betting markets.

KEYWORDS: Market Efficiency, Sports Analytics, Sports Betting

INTRODUCTION

Sports analytics have become ubiquitous in print, online and on television. Media outlets such as ESPN and Five Thirty Eight have a large analytics staff devoted to developing, refining, and communicating analytics-based predictions. ESPN now regularly provides a win probability for every game in major sports on their chyron generally labelled *according to ESPN Analytics*. The popular website FiveThirtyEight.com, well known for predictive modelling in politics as well as sports, publishes several models that provide forecasted win probabilities in major sports for every game.

Both the ESPN model and the FiveThirtyEight models are variations of a Power Index model. While each model has unique features, they are very similar in structure. Each team has a power ranking that indicates its overall capability. The probability that a team defeats another team is based on the relative values of their power rankings. Based on the outcome of a game power ranking points are taken from the loser and given to the winner, thus over the course of the season power ranking points migrate from poor teams to good teams.

Given the significant level of resources these organizations devote to their modeling it is reasonable to ask how well these model predictions perform. In this paper we will restrict ourselves to the National Football League (NFL). We will review the ESPN Football Power Index (ESPN FPI) and the two FiveThirtyEight Elo based models, the Elo model and the QB Elo model. We will discuss the concepts behind the models and review, to the extent possible, how they work. We will then evaluate the accuracy of these models in a probabilistic sense, comparing them to actual results and each other. Finally, we will compare these models to the probabilistic forecast provided by betting odds. We will test the model's ability to generate positive betting returns by following their recommendations, and in the process test if these models threaten the semi-strong version of the efficient market hypothesis as applied to the betting markets. We will then discuss what, if anything, they tell us about semi-strong efficiency overall.

LITERATURE REVIEW

With the increased application of analytics to various fields of study there is a reasonably large body of research on the use of models to predict sports outcomes. Horvat and Job (2020) provide a detailed review of the application of machine learning to sports prediction. Their analysis covers over 100 papers that cover multiple sports from Cricket to American Football. Some papers address models across sports (Bunker & Thabtah, 2019) while other papers address individual sports such as basketball (Huang & Lin, 2020; Osken & Onay, 2022; Zhao, Du, & Tan, 2023), hockey (Pischedda, 2014), baseball (Huang & Li, 2021; Soto-Valero, 2016), soccer (Dandıl & Ergez, 2018; Hassan, Akl, Hassan, & Sunderland, 2020; Mattera, 2021), collegiate wrestling (Bigsby & Ohlmann, 2017), golf (Chae, Park, & So, 2021), rugby (M. Bennett, Bezodis, Shearer, & Kilduff, 2020) cricket (Balasundaram, Dhandapani, & Ashokkumar, 2022; Patil, Duraphe, Motarwar, Suganya, & Mariappan, 2023; Shakil, Abdullah, Momen, & Mohammed, 2020) and tennis (Wilkens, 2021).

Some papers explicitly consider betting in certain sports such as college football (R. Bennett, 2020; Coleman, 2015) Predictive analytics are very popular in the NFL, with outsiders and insiders. The NFL has itself partnered with organizations such as Amazon Web Services (AWS) to exploit machine learning technology (Learning, 2021). Predictive models for in game win probabilitiy assessment and player evaluation have been published in the lierature and made available as libraries for the popular statistical package R (Yurko, Ventura, & Horowitz, 2019).

Capital Market Efficiency

Our analysis includes an assessment of the efficiency of the sports betting markets. Market Efficiency is a concept first developed in the economics and finance literature as defined by the efficient market hypothesis. (EMH). The efficient-market hypothesis (EMH) is a hypothesis in financial economics that states that asset prices *fully reflect* all available information. A direct implication is that it is impossible to "beat the market" consistently on a risk-adjusted basis since market prices should only react to new information (Wikipedia, 2022).

A key review of the theoretical and empirical literature on the Empirical Market Hypothesis is provided in Fama (1970). Fama analyzes efficient markets relative to three board information sets. Weak form efficiency is based on the use of historical prices, semi-strong efficiency on publicly available information, and strong efficiency is based on all information public and private. In this framework the possibility of trading systems based on the relevant information set generating excess returns, that is returns in excess of equilibrium expected profits, are ruled out. Simply put, excess profits are not possible from a trading system in a market that is efficient. The empirical analysis in this paper provides reasonably strong support for the weak, and semi-strong forms of market efficiency. The analysis, however, identified exceptions to the strong form of market efficiency whereby market makers and corporate insiders can exploit their monopolistic access to information in order to earn returns in excess of the expected risk-adjusted rate.

Sports Betting Market Efficiency

A large body of research published beginning in the 1980s and 90s examined efficiency in sports betting markets. Much of this early research focused on pari-mutual and fixed odds systems in horse racing. A comprehensive review of this literature is provided in (Kuypers, 2000). Kuypers reviews five papers that examine pari-mutual systems, six papers on odds-based systems, and four papers on spread based systems. Seven of the reviewed papers assess weak form efficiency, while six examine semi-strong efficiency and two assess strong-form efficiency.

Kuypers uses the following definitions of efficiency in the sports betting context:

- Weak form: no abnormal returns, either to the bookmaker or the bettor, can be achieved solely from price information. An abnormal return is defined as a return different from the bookmaker's expected take.
- Semi-strong: no abnormal returns can be achieved from odds or any publicly available information.
- Strong: no abnormal returns can be achieved by any group in society incorporating odds publicly available and privately available information.

A reasonably large body of literature examines various aspects of betting efficiency across different sports and different betting markets. The majority of these papers focus on weak form efficiency by testing startaegies based on polint spreads. Gray and Gray (1997) examine efficiency in NFL point-spread markets using a probit model approach to see if the spread is an unbiased predictor. They test various betting strategies including betting on teams where their model forecasts a greater than 52.4% win probability; the implied win probability in most point-spread bets. Efficiency tests for NFL betting have been the subject of several economics thesis projects (Anderson, 2019; Hetherington, 2006; Williams, 2021) Hetherington (2006) looks at NFL prediction markets with a behavioral finance perspective. Stetzka and Winter (2021) investigate the rationality of gambling overall. Anderson (2019) uses OLS regression to evaluate various strategies for betting against the spread in NFL games. (Williams, 2021) uses an econometric approach to analyze rationality and efficiency. The analysis finds evidence of bias but cannot show any market inefficiency.

Robbins (2023) examines weak form efficiency across moneyline bets on multiple North American professional and collegiate sports. He finds technical violations of market efficiency including longshot biases and differential returns on various odds ranges, but nothing that leads to a long term consistent positive return. Arscott (2023) examines efficiency in college football betting.

Most of these analyses are focused on weak-form efficiency. In sports betting weak form efficiency implies that decision rules based simply on the stated odds will not earn a long-term profit, and all odds ranges will earn a similar negative return. Semi-strong efficiency implies that profitable strategies cannot be developed based on publicly available information. One implication of semi-strong efficiency in sports betting is that an analytical model that uses publicly available information will not identify bets that generate a positive long-term return.

BETTING ODDS AND PROBABILITIES

The menu of bets that can be made on sports is very large. Bets can be placed on almost anything related to a game, a team, or even individual performances of players. With minor variations from sport to sport, the main betting options have three different components: totals, spreads and moneyline.

- Totals: the total, or over/under, is a bet on the total points scored in the game. Bettors can bet the total points will be over, or under the stated line.
- Spreads: a bet on a team to win by a certain margin. The underdog is bet with plus points, the favored with negative points.

- Moneyline: a straight bet on what team will win the game. Moneyline bets are made with differential payouts such that a bet on a favorite will risk more than can be won, while a bet on an underdog will return more than the amount risked.

Note that both totals and spread bets are quoted along with moneyline odds so that the payout to a winner is less than the amount risked. Odds are stated in different equivalent formats in different locations and different settings. In the United States odds are most often quoted in American Odds format. Our analysis will focus on moneyline bets for the NFL.

In the American format the odds can be expressed as either a positive number or a negative number. A positive number shows the profit a successful wager will return on a $100 bet. For example, a bettor who wagers $100 at +110 odds and wins, will earn a profit of $110, plus the original wager of $100 for a total payout of $210. Positive odds typically imply the team is an underdog. Conversely, negative odds show how much a bettor must risk to earn a $100 profit. For example, if a bet is made for $120 at -120 odds, the successful bettor will receive a profit of $100, plus the original wager of $120 for a total payout of $220. The favorite team is given negative odds, but in some evenly matched games both teams may have negative odds. More formally the Payout *P* to a wager of stake *S*, at odds *M* are given by equation (1).

$$P = \begin{cases} S \times \dfrac{M}{100} + S & \text{for M} > 0 \\ \dfrac{S}{-M/100} + S & \text{for M} \leq 0 \end{cases} \tag{1}$$

Odds of +100- and -100 are equivalent. In practice *M* is always quoted as a number with an absolute value greater than or equal to 100. So, while odds of -125 and +80 would both return a profit of $80on a $100 bet, the odds are always quoted as -125.

Moneyline odds carry an implied probability of success. The implied probability is the probability at which a bettor is indifferent to taking either side of the bet. The probability calculation in the American odds format again depends on whether the odds are positive or negative. So, for a bet with odds *M*, the implied probability *p* is given by

$$p = \begin{cases} \dfrac{100}{M + 100} & \text{for M} > 0 \\ \dfrac{-M}{-M + 100} & \text{for M} \leq 0 \end{cases} \tag{2}$$

While equation (2) gives the odds on one side of a bet, the bookmaker quotes odds in pairs. For example, a bookmaker might quote odds of -120 for a favorite and +110 for the underdog. Converting each of these to implied probabilities gives probabilities of 54.5% and 47.6%. These odds are not fair, in the sense that they add up to more than 100%. The excess probability, in this example 2.1%, is the book margin (*k*), sometimes referred to as the *vig* or the *juice*. The book margin exists so that the bookmaker is guaranteed a profit as long as bets are made in the appropriate proportion. Book margins in the range of 3%-5% are common.

In order to convert the bookmaker's odds into meaningful probability estimates the odds must be converted to consistent probabilities. Draws are rare in football. So, if the contest ends in a draw all win-lose bets are effectively cancelled, and bettors are returned their original stake. The most common way to convert the implied probabilities is a simple normalization process.

So, for a contest with implied probabilities of p_1 and p_2, the normalized probability that team 1 will win the game and bets will pay is

$$p_{1_n} = \frac{p_1}{p_1 + p_2} \tag{3}$$

<u>The Sportsbook's Margin</u>
Because the implied odds are unfair, they add up to more than one, the sportsbook has a built-in advantage. The excess probability gives the sportsbook a built-in margin, appropriately allocated bets on either side will guarantee the book a profit. The sportsbook's profit margin is proportional to the book sum, the excess implied probability in the stated odds. If we have a two-way bet with implied odds p_1 and p_2, then the book sum k, is given by

$$k = p_1 + p_2 - 1 \tag{4}$$

The bookmaker's margin (m), also known as the *hold*, is the sportsbook's average profit and can be shown to be

$$m = \frac{k}{k+1} \tag{5}$$

<u>The Bettor's Decision</u>
The bettor is presented with a series of potential bets and must decide which bets to accept. Assume for simplicity and without loss of generality that the odds for a particular bet are presented at positive American Odds M and the bettor will make a bet of one dollar.

The expected profit from the bettor's perspective is

$$E[P] = p_b \frac{M}{100} - (1 - p_b) \tag{6}$$

Where p_b is the bettor's subjective probability assessment that his bet will win. The bettor's expected profit is positive if

$$p_b > \frac{100}{M + 100} \tag{7}$$

The right-hand side of equation (7) is the implied probability of the odds as expressed in equation (2). The decision of the bettor is then quite simple, accept bets where you assess the probability of victory to be greater than the probability implied by the odds. It is important to note that does not mean the bettor is simply choosing sides on every bet. Because the bookmaker is offering unfair odds there will be a number of bets where neither side is assessed as a positive EV bet. This is especially true if the bettor's odds assessments and those of the market, as expressed by the bookmaker are similar to each other.

FOOTBALL POWER INDEX

The ESPN predictive model is known as the ESPN Football Power Index (ESPN FPI). The ESPN FPI is a rating system developed by ESPN that measures team strength and uses it to forecast game and season results (ESPN). The FPI gives each team a rating number that indicates their strength. The relative strengths of two teams competing against each other are then used to determine the win probability for each team. Based on the results of each game team strength ratings are updated.

The models used by FiveThirtyEight.com are methodologically very similar to ESPN's model. But while ESPN treats the details of their model as proprietary, FiveThirtyEight is generally fairly transparent about how their models work (Silver, 2022). FiveThirtyEight introduced their model, which they called NFL Elo Ratings in 2014 (Silver, 2014) and modified the approach somewhat in 2015 (Paine, 2015). Starting in 2019 FiveThirtyEight introduced a second model similar to the original along with a quarterback adjustment, the QB Elo model. FiveThirtyEight describes their models at a reasonable level of detail and provides access to the data used to in the model on Github (FiveThirtyEight.com). The FiveThirtyEight models are variations on the Elo model, originally developed by Arpad Elo to rank chess players (Wikipdeia). The application of Elo models to sports is described in Aldous (2017).

FiveThirtyEight Elo Model

The basic Elo model was developed by Arpad Elo as a tool for ranking chess players, but has since been adapted to sports, including the NFL. The Elo model assigns each team a numerical ranking. In the NFL the Elo ranking is scaled so that the average team has a ranking of 1505 points. For the 2022 season ranking ranged from a low of 1291 to a high of 1719.

The forecast for each game is expressed as a win probability. The base probability that team *A*, with Elo score R_A will defeat team *B*, with Elo score R_B is given as

$$\Pr(A) = \frac{1}{10^{\frac{R_B - R_A}{400}} + 1} \qquad (8)$$

The pre-game Elo point differential is adjusted for several factors including a home field advantage of 4 points per thousand miles travelled buy the visitor, and a 25-point rest adjustment for teams coming off a bye week. In the QB Elo model an additional adjustment is made based on the quarterback.

After the game Elo points are adjusted based on the outcome, in a zero-sum exchange where the winner takes points from the loser. The base level of points transferred is derived based on how likely the team was to win based on the pregame Elo ratings. If E_A is the probability team *A* would win, S_A is 1 if team *A* won, 0 otherwise and *K* is a scaling constant, Then the new Elo rating for team *A* is

$$R_A^{'} = R_A + K(S_A - E_A) \qquad (9)$$

K is a factor that determines how quickly ratings react to new results. The larger the *K* value the more points shift. The *K* value for football is approximately 20. Sports with more games, where each individual game has less impact, like baseball or basketball, will have smaller *K* values.

Elo scores are also impacted by the margin of victory (winner's points-loser's points). The effect is implemented as a multiplier. The larger the margin, the larger the multiplier. But the number of points has a declining marginal benefit.

$$MovMultiplier = \ln(WinnerPointDiff + 1) \times \frac{2.2}{WinnerELODiff \times 0.001 + 2.2} \qquad (10)$$

Overall, the shift in Elo points per game is relatively small. During the 2022 NFL season the median point shift was 17.13 points.

The ESPN FPI Model
ESPN is significantly less transparent about their model than FiveThirtyEight, providing only general descriptions of the model and the model results. They recently published an overview of the model along with pre-season ratings for the 2023 season (Walder, 2023). Like the FiveThirtyEight model the FPI gives each team a Power Ranking Score. Win probabilities are calculated based on the difference in ranking score. Team rankings are updated during the season based on opponent score, game location, and point differential. The FPI model differs from FiveThirtyEight in that it accounts for the separate impact of offense, defense, and special teams based on the Expected Points Added (EPA) metric. The ESPN model also differs in how it determines pre-season rankings, with the ESPN model utilizing betting odds as an initial factor.

MODEL ACCURACY

The most basic assessment of a predictive model such as the FPI models is how often they are right, i.e., what proportion of the teams the model favors to win actually end up winning. A standard baseline for comparison is the coin-flip which would over the long term be accurate 50% of the time. However, winning percentages in most major sports, including football, are biased in favor of the home team. The home team in the NFL wins about 55% of the time, so simply betting on the home team will result in roughly a 55% win rate. But recall that since bets are not even money, and the home field advantage is built into the odds, winning 55% will not necessarily provide a profit. We will return to this issue in the next section, but for now we compare the winning percentage of the three models with the home field winning percentage. We also consider a second benchmark, the winning percentage of the team favored by the betting odds. Those complete results are presented in Table 1.

Table 1-Winning Percentages

Accuracy Picking Game Winner						
Season	Games	Home	ELO	QBELO	FPI	Odds
2014	267	57.5%	69.2%	-	-	66.2%
2015	267	53.2%	65.9%	-	62.5%	63.3%
2016	267	58.5%	64.5%	-	62.3%	65.3%
2017	267	56.9%	65.2%	-	67.2%	68.9%
2018	267	59.6%	62.6%	-	64.2%	64.5%
2019	267	51.9%	63.9%	64.3%	62.4%	64.3%
2020	269	49.6%	65.3%	68.3%	66.7%	67.2%
2021	284	50.9%	59.0%	60.4%	60.1%	62.2%
2022	271	55.4%	62.5%	62.1%	59.9%	64.7%
Total	2,426	54.8%	64.2%	63.7%	63.1%	65.1%

The home team won a majority of games in every year except 2020, which was the Covid year where attendance restrictions muted the home field advantage. Each of the three models consistently predicted winners with a higher accuracy rate than the naïve home field rule and often by reasonably large margins. The models were, however, generally less accurate than the betting odds. The Elo model outperformed the betting odds in its first two years, but not since. Similarly, the QB Elo model did as well or better than the betting odds in its first two years, but not since. In the aggregate none of the models has been as accurate as the betting odds.

Accuracy is an interesting, but insufficient metric for evaluating the betting utility of the models. It depends how the model does at various odds levels. To begin assessing this we present calibration graphs. To create the calibration graph we divide the predictions into quintiles, twenty evenly sized groups based on the win probability implied by the model. For each quintile we calculate the average implied win probability and the actual win probability. The graph shows these odds in a scatter plot format. If the odds are well calibrated, teams with a 30%-win probability will win 30% of the time and the points on the calibration chart will align on the diagonal.

Figure 1 is the calibration plot for the base Elo model. The model appears to be reasonably well calibrated as none of the points diverge from the diagonal by much. However, the model does appear to exhibit some bias. Of the 10 points with forecasted probabilities less than 50%, 8 are above the line. Conversely points with win probabilities above 50% are generally below the line. The implication is that the Elo forecast is biased. It tends to undervalue underdogs, that is give them a higher lower probability than they achieve and overvalue favorites. Figure 2 shows the corresponding graph for the QB Elo model. The points here diverge from the line more than the Elo model, and the same general bias is evident; underdogs mostly overperform and favorites mostly underperform. Finally, Figure 3 has the calibration chart for the ESPN FPI model. It exhibits a similar pattern to the Elo models.

Figure 1- Elo Win Probability Calibration

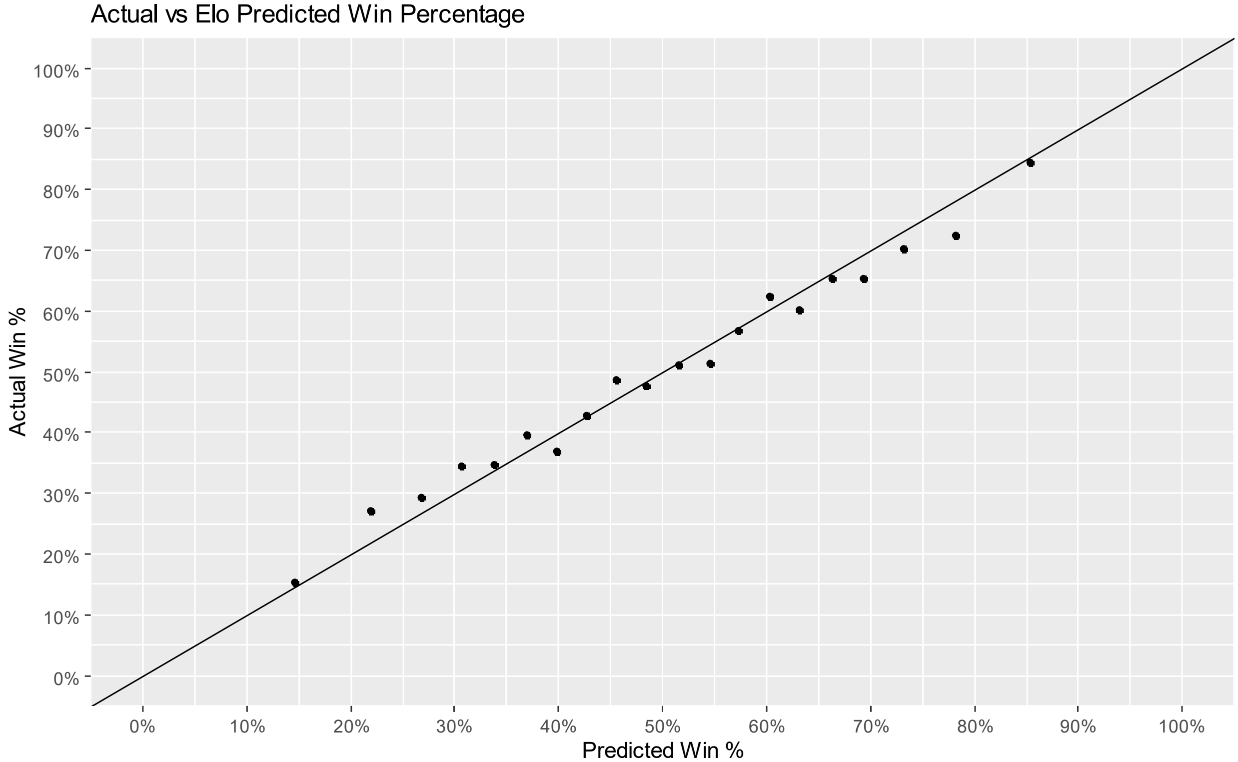


Figure 2-QB Elo Win Probability Calibration

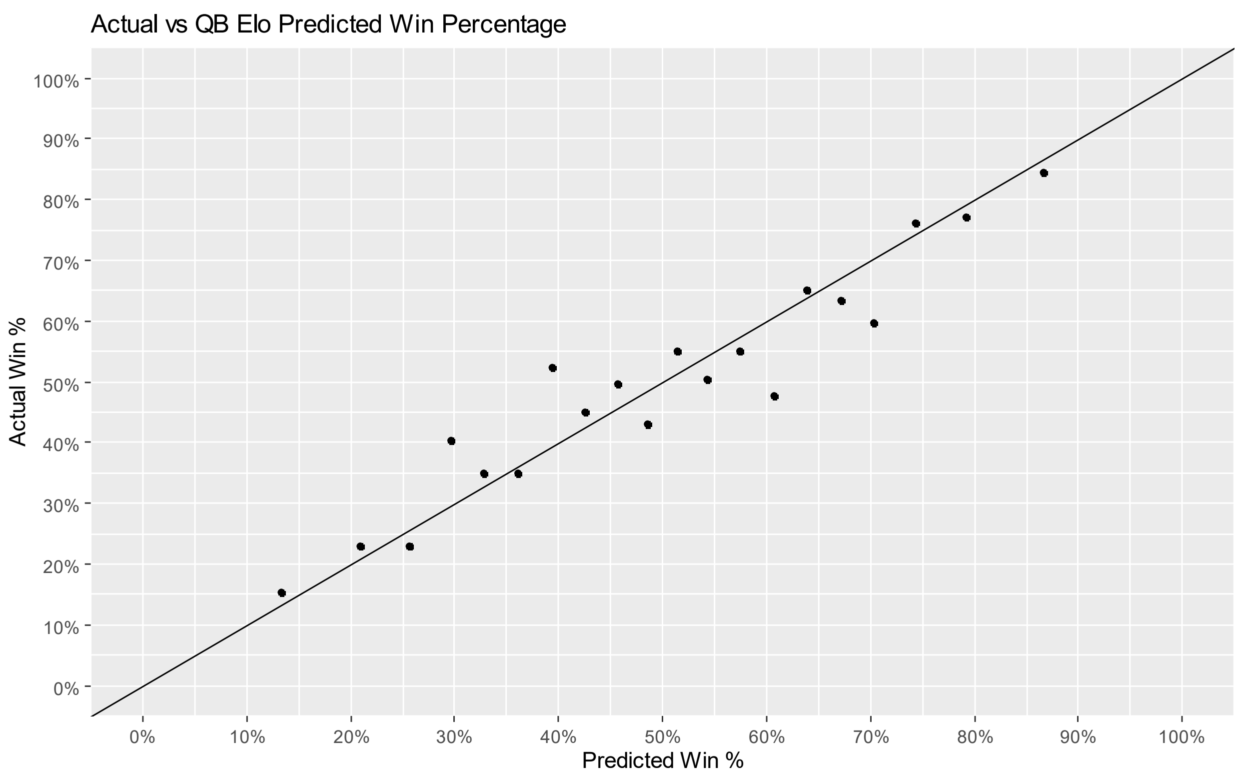
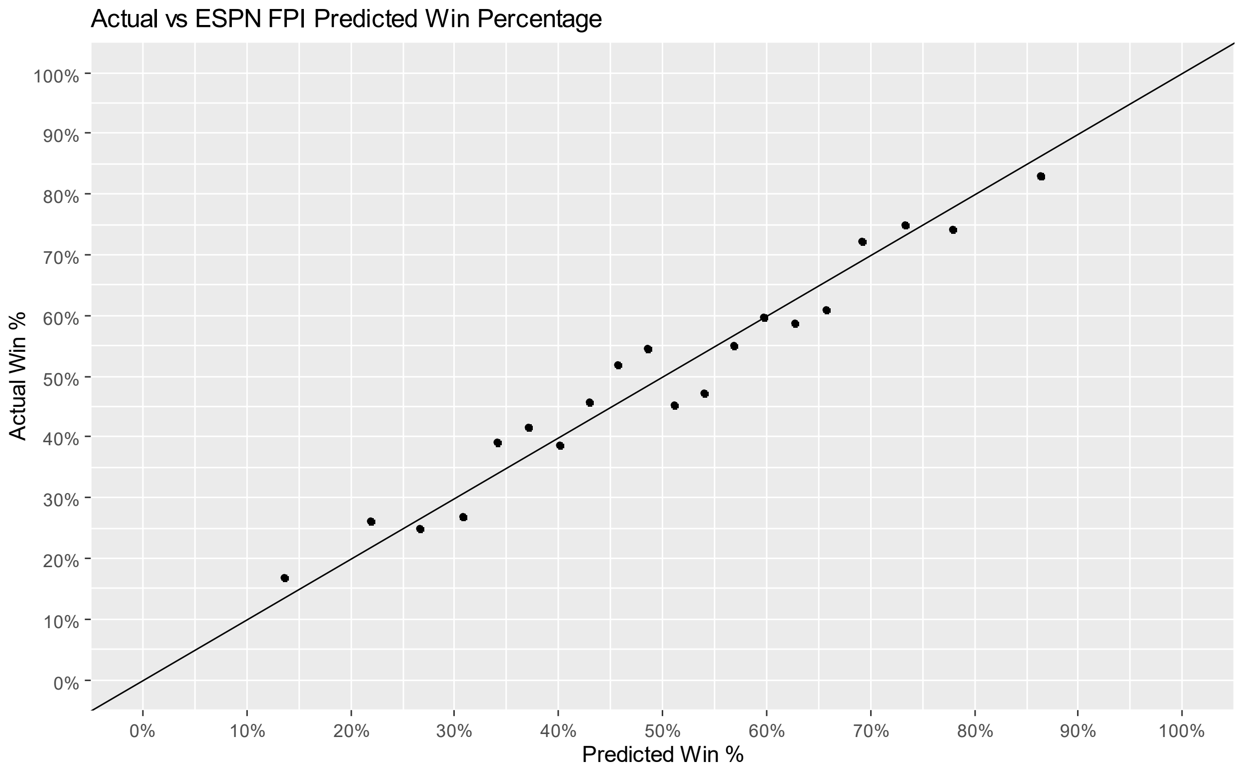


Figure 3-ESPN FPI Win Probability Calibration



The graphical information presented in these figures is summarized numerically in Table 2. From here we can verify that of the ten underdog bins the actual win rate was higher in eight bins for the Elo model, 7 for the AB Elo model, and seven for the ESPN FPI model. This confirms the observation that these models are biased in the sense that they undervalue the under dogs.

To get a sense of the overall accuracy of the probabilistic predictions we calculate the Brier score and show the results by season in Table 3. The Brier score is a metric commonly used to evaluate the accuracy of classification models. For a unidimensional model such as a win-loss model the Brier score is equivalent to the mean squared error of the prediction. The Brier score (B) is given by equation (11), where *f* is the forecasted probability of victory and *o* is an indicator variable that is 1 if the team wins and 0 otherwise, and *N* is the number of predictions made, which is two times the number of games.

$$B = \frac{1}{N}\sum_{i=1}^{n}(f_t - o_t)^2 \tag{11}$$

A naïve predictor that gives each team a 50% chance of winning will have a Brier score of .25. A perfect Brier score of 0 is possible, but only if the model assigns 100% probability to every prediction and is always correct. So, practically we would hope to see Brier scores below .25. Each of the models outperform the naïve forecast and register Brier scores in the .207 to .235 range. The best performance was the Elo model in its first year, but that performance tended to decline over time rising from .207 in 2014 to .235 in 2022. In the most recent 4 years when all models have been making projections, the ESPN model had the best score in two years (2019 and 2021) and the worst score in 2022. The enhanced Elo model, the QB Elo, outperformed the base Elo model in all 4 years.

Table 2- Win Probabilities

Forecasted and Actual Win Rates by Quintile

For Games with Predictions

	538 Elo			538 QB Elo			ESPN FPI		
Bin	ELO_N	ELO_PWP	ELO_AWP	QBELO_N	QBELO_PWP	QBELO_AWP	ESPN_N	ESPN_PWP	ESPN_AWP
1	243	14.7%	15.2%	110	13.4%	15.5%	215	13.7%	16.7%
2	243	21.9%	27.2%	109	20.9%	22.9%	215	22.0%	26.0%
3	242	26.8%	29.3%	109	25.6%	22.9%	222	26.6%	24.8%
4	243	30.7%	34.6%	109	29.7%	40.4%	213	30.8%	26.8%
5	242	33.7%	34.7%	109	32.8%	34.9%	215	34.0%	39.1%
6	243	36.9%	39.5%	109	36.1%	34.9%	212	37.1%	41.5%
7	242	39.8%	36.8%	109	39.3%	52.3%	218	40.1%	38.5%
8	243	42.7%	42.8%	109	42.5%	45.0%	212	42.9%	45.8%
9	242	45.5%	48.8%	109	45.7%	49.5%	214	45.7%	51.9%
10	243	48.4%	47.7%	109	48.6%	43.1%	218	48.5%	54.6%
11	243	51.6%	51.0%	109	51.4%	55.0%	217	51.2%	45.2%
12	242	54.5%	51.2%	109	54.3%	50.5%	212	54.0%	47.2%
13	243	57.3%	56.8%	109	57.5%	55.0%	218	56.8%	55.0%
14	242	60.2%	62.4%	109	60.7%	47.7%	216	59.7%	59.7%
15	243	63.1%	60.1%	109	63.9%	65.1%	211	62.7%	58.8%
16	242	66.3%	65.3%	109	67.2%	63.3%	213	65.8%	61.0%
17	243	69.3%	65.4%	109	70.3%	59.6%	216	69.1%	72.2%
18	242	73.2%	70.2%	109	74.4%	76.1%	214	73.3%	74.8%
19	243	78.1%	72.4%	109	79.1%	77.1%	217	77.9%	74.2%
20	243	85.3%	84.4%	110	86.6%	84.5%	212	86.3%	83.0%

Table 3- Brier Scores

Season by Season Brier Scores For Games with Predictions			
Season	ELOBrier	QBELOBrier	ESPNBrier
2014	0.207	-	-
2015	0.225	-	0.227
2016	0.219	-	0.221
2017	0.217	-	0.215
2018	0.223	-	0.210
2019	0.225	0.220	0.218
2020	0.219	0.209	0.210
2021	0.235	0.232	0.229
2022	0.226	0.221	0.227

MODEL PREDICTIONS VS BETTING ODDS

The 3 forecasting models outperform the naïve 50-50 model as well as the slightly less naïve model of picking the home team, but how do they do as compared to the betting markets. To evaluate this, we implement a simple betting rule, bet all teams where the model indicates a win probability in excess of the implied probability of the betting odds. This is the definition of a positive expected value bet we outlined in equation (7).

It is important to note that this decision rule will not imply a bet on every game. Because the bookmaker is offering unfair odds, the model must forecast a win probability a few points higher than the betting odds. Stated another way, if the betting odds and the forecasting model assess the same win probability our decision rule would not bet either side.

As a baseline we also test the naïve approach of betting on the home team when odds are available. The results of this analysis are presented in Table 4. This analysis assumes that for each game where the decision rule indicates a bet that $100 is wagered. For each category the table provides the aggregate dollar impact with the average profit per bet. Since each bet is $100, the average represents the percentage return to bets.

Table 4- Results from Rule Based Betting

Profit from Picking Based on Model Forecasts												
Season	Bets	TotalProfit	HomeAvgProfit	ELOBets	TotalELOProfit	ELOPickAvgProfit	QBELOBets	TotalQBELOProfit	QBELOPickAvgProfit	ESPNBets	ESPNTotalProfit	ESPNPickAvgProfit
2014	263	−782	−2.97	217	1,144	5.27	-	-	-	-	-	-
2015	263	−1,915	−7.28	222	1,687	7.60	-	-	-	191	3,384	17.72
2016	265	−679	−2.56	229	−95	−0.42	-	-	-	217	−1,086	−5.01
2017	266	−1,157	−4.35	224	−1,548	−6.91	-	-	-	196	−1,035	−5.28
2018	266	59	0.22	227	−284	−1.25	-	-	-	195	1,477	7.57
2019	264	−3,316	−12.56	217	−3,036	−13.99	215	−1,906	−8.87	197	784	3.98
2020	268	−3,639	−13.58	227	−2,602	−11.46	208	−237	−1.14	207	−5,072	−24.50
2021	280	−1,460	−5.21	245	−858	−3.50	227	−1,917	−8.45	230	11	0.05
2022	267	−354	−1.33	232	−1,977	−8.52	213	−1,728	−8.11	227	−2,668	−11.75
Total	2,402	−13,244	−5.51	2040	−7,571	−3.71	863	−5,788	−6.71	1,660	−4,205	−2.53

The table shows that betting on the home team is generally a poor bet. It has a negative return in all years except for 2018 where it has a very small positive return of 0.22%. Recall that the home team wins a majority of games in most years, but that fact is crucially reflected in the betting odds. The overall loss rate of 5.5% is roughly in line with the expected profit of the bookmaker based on the hold they apply to their odds.

The Elo model makes predictions that differ enough from the betting odds to wager on a large number of games, roughly 85% of the games. The Elo model outperforms the naïve model but over the course of the data set earns a negative return rate of –3.7%. Interestingly the Elo model showed positive results in its first two years and then turned increasingly negative. The QB Elo model which debuted later is somewhat more selective in the number of games where its prediction warrants a bet. Its performance against the odds was better than Elo in three out of four years, but over its 4-year history it earns a negative -6.7% return. It significantly underperformed the naïve rule in the last two seasons.

The ESPN FPI model has the most volatile returns of the three models. Its reported results in 2014, the first year that predictions are available from the ESPN website are extremely positive at +17.7%. But in 2020 it recorded a -24.5% return, far worse than the home field naïve rule. Overall, the ESPN model has the best returns of the three models, but still comes in at -2.5% over the 8-year history.

CONCLUSION

Both ESPN and FiveThirtyEight have significant resources dedicated to analytics across multiple sports, and in the case of FiveThirtyEight across multiple fields. ESPN regularly and actively promotes their individual game predictions on their network. Their models are reasonably accurate, picking the winner as the favorite more than 60% of the time. The models are reasonably well calibrated though it appears that each model is biased in that they tend to underrate underdogs.

While these models are reasonably accurate, none of them can outperform the betting odds on any consistent basis. They all earn negative returns over the long term only moderately better than the naïve rule of betting the home team. These models are reasonably sophisticated, with access to large data sets, both broad in scope and long in history. Betting market odds are in contrast largely the result of the un-coordinated action of many bettors. While sports books may also use analytics to set initial odds, they adjust based on market forces. The betting odds represent a market-based consensus, and the odds makers have the undeniable and very significant advantage of setting unfair odds and implementing a hold. That being said, even ignoring the hold and utilizing the normalized prediction the odds are more accurate than the models.

So, does this analysis tell us anything about market efficiency in the semi-strong format. Remember that if betting markets are semi-strong efficient, no analytical model will generate positive returns. Our analysis shows that at least so far, none of the three models we analyzed generate positive returns. That does not prove that the betting markets are efficient, but it fails to show that they are not. Had we been able to find a consistent positive return from any of these models we could declare the market inefficient. But we have not done so.

It could well be the case that some highly skilled modeler somewhere has developed a model that earns long term positive returns but has kept the model details private for selfish reasons. The real implication of market efficiency is that publicly available information is quickly incorporated into prices by market mechanisms. The implications of this theory are that no matter how good a model such as the ESPN Football Power Index, or any variation of the Elo model becomes, those predictions would be quickly incorporated into the odds pricing out any long-term profit. Since the models we have discussed are so public efficient market theory implies their predictions are incorporated into the betting off their predictions will not earn a positive return. While it is highly speculative, the fact that model predictions, in particular the Elo and ESPN models, had their best results in the first year supports the notion that once these models came into the public domain their predictions were priced in as efficient market theory suggests.

REFERENCES

Aldous, D. (2017). Elo Ratings and the Sports Model: A Neglected Topic in Applied Probability? *Statistical Science, 32*(4), 616-629. Retrieved from http://www.jstor.org/stable/26408861

Anderson, S. (2019). *NFL Betting Market Efficiency: Finding a Profitable Betting Strategy.* Skidmore College, Retrieved from https://creativematter.skidmore.edu/econ_studt_schol/121/

Arscott, R. (2023). Market Efficiency and Censoring Bias in College Football Gambling. *Journal of Sports Economics, 24*(5), 664-689. doi:10.1177/15270025221148991

Balasundaram, A., Dhandapani, K., & Ashokkumar, S. (2022). *Predicting different facets in the game of cricket using machine learning* (Vol. 2418).

Bennett, M., Bezodis, N., Shearer, D., & Kilduff, L. (2020). Predicting performance at the group-phase and knockout-phase of the 2015 Rugby World Cup. *European Journal of Sport Science, 21*, 1-18. doi:10.1080/17461391.2020.1743764

Bennett, R. (2020). Next game reaction to mispriced betting lines in college football. *Applied Economics Letters, 28*, 1-4. doi:10.1080/13504851.2020.1795063

Bigsby, K. G., & Ohlmann, J. W. (2017). Ranking and prediction of collegiate wrestling. *Journal of Sports Analytics,, Forthcoming*.

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics, 15*(1), 27-33. doi:https://doi.org/10.1016/j.aci.2017.09.005

Chae, J., Park, J., & So, W.-Y. (2021). Victory Prediction of Ladies Professional Golf Association Players: Influential Factors and Comparison of Prediction Models. *Journal of Human Kinetics, 77*, 245-259. doi:10.2478/hukin-2021-0023

Coleman, B. (2015). Team Travel Effects and the College Football Betting Market. *Journal of Sports Economics*. doi:10.1177/1527002515574514

Dandıl, E., & Ergez, C. (2018). *Prediction of Soccer Results Using Artificial Neural Networks with Web-Based Automation*.

ESPN. Breaking Down the Football Power Index. Retrieved from https://www.espn.com/video/clip/_/id/20457171

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*(2), 383-417. doi:10.2307/2325486

FiveThirtyEight.com. NFL Elo. Retrieved from https://github.com/fivethirtyeight/data/tree/master/nfl-elo

Gray, P. K., & Gray, S. F. (1997). Testing Market Efficiency: Evidence From The NFL Sports Betting Market. *The Journal of Finance, 52*(4), 1725-1737. doi:https://doi.org/10.1111/j.1540-6261.1997.tb01129.x

Hassan, A., Akl, A.-R., Hassan, I., & Sunderland, C. (2020). Predicting Wins, Losses and Attributes' Sensitivities in the Soccer World Cup 2018 Using Neural Network Analysis. *Sensors, 20*, 3213. doi:10.3390/s20113213

Hetherington, R. A. (2006). *Hetherington, R. Alexander, Betting Against Efficiency: Behavioral Finance in an Nfl Gambling Exchange (March 26, 2006). Available at SSRN: https://ssrn.com/abstract=881514 or http://dx.doi.org/10.2139/ssrn.881514*. SSRN.

Horvat, T., & Job, J. (2020). The use of machine learning in sport outcome prediction: A review. *WIREs Data Mining and Knowledge Discovery, 10*(5), e1380. doi:https://doi.org/10.1002/widm.1380

Huang, M.-L., & Li, Y.-Z. (2021). Use of Machine Learning and Deep Learning to Predict the Outcomes of Major League Baseball Matches. *Applied Sciences, 11*, 4499. doi:10.3390/app11104499

Huang, M.-L., & Lin, Y.-J. (2020). Regression Tree Model for Predicting Game Scores for the Golden State Warriors in the National Basketball Association. *Symmetry, 12*, 835. doi:10.3390/sym12050835

Kuypers, T. (2000). Information and Efficiency: An Empirical Study of a Fixed Odds Betting Market. *Applied Economics, 32*, 1353-1363. doi:10.1080/00036840050151449

Learning, A. M. (2021). *The NFL "goes long" on machine learning*. Retrieved from https://pages.awscloud.com/hk-machine-learning-nfl.html

Mattera, R. (2021). Forecasting binary outcomes in soccer. *Annals of Operations Research*, 1-20. doi:10.1007/s10479-021-04224-8

Osken, C., & Onay, C. (2022). Predicting the winning team in basketball: A novel approach. *Heliyon, 8*, e12189. doi:10.1016/j.heliyon.2022.e12189

Paine, N. (2015). NFL Elo Ratings Are Back! Retrieved from https://fivethirtyeight.com/features/nfl-elo-ratings-are-back/

Patil, R., Duraphe, A., Motarwar, P., Suganya, G., & Mariappan, P. (2023). Cluster-Centric Based Hybrid Approach for Cricket Sports Analytics Using Machine Learning. In (pp. 268-283).
Pischedda, G. (2014). Predicting NHL Match Outcomes with ML Models. *International Journal of Computer Applications, 101*, 15-22. doi:10.5120/17714-8249
Robbins, T. R. (2023). Weak Form Efficiency in Sports Betting Markets. *American Journal of Management, 23*(2). doi:https://doi.org/10.33423/ajm.v23i2.6051
Shakil, F., Abdullah, A., Momen, S., & Mohammed, N. (2020). Predicting the Result of a Cricket Match by Applying Data Mining Techniques. In (pp. 758-770).
Silver, N. (2014). Introducing NFL Elo Ratings. Retrieved from https://fivethirtyeight.com/features/introducing-nfl-elo-ratings/
Silver, N. (2022). How Our NFL Predictions Work. Retrieved from https://fivethirtyeight.com/methodology/how-our-nfl-predictions-work/
Soto-Valero, C. (2016). Predicting Win-Loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport, 15*, 91-112. doi:10.1515/ijcss-2016-0007
Stetzka, R. M., & Winter, S. (2021). How rational is gambling? *Journal of Economic Surveys, n/a*(n/a). doi:https://doi.org/10.1111/joes.12473
Walder, S. (2023). NFL Football Power Index: 2023 projections, Super Bowl chances. Retrieved from https://www.espn.com/nfl/story/_/id/37681787/nfl-football-power-index-2023-season-projections-super-bowl-chances-playoff-draft
Wikipdeia. Elo rating system. Retrieved from https://en.wikipedia.org/wiki/Elo_rating_system
Wikipedia. (2022). Efficient-market hypothesis. Retrieved from https://en.wikipedia.org/wiki/Efficient-market_hypothesis
Wilkens, S. (2021). Sports prediction and betting models in the machine learning age: The case of tennis. *Journal of Sports Analytics, 7*, 99-117. doi:10.3233/JSA-200463
Williams, A. (2021). *Rationality and Efficiency in NFL Gambling Markets.* Ohio State University, Retrieved from https://kb.osu.edu/handle/1811/92881
Yurko, R., Ventura, S., & Horowitz, M. (2019). nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports, 15*(3), 163-183. doi:doi:10.1515/jqas-2018-0010
Zhao, K., Du, C., & Tan, G. (2023). Enhancing Basketball Game Outcome Prediction through Fused Graph Convolutional Networks and Random Forest Algorithm. *Entropy, 25*, 765. doi:10.3390/e25050765

DECISION SCIENCES INSTITUTE

Lean Manufacturing Practices Impacting Organizational Competitiveness Through Better Alignment of Benchmarking Performance Metrics

(Full Paper Submission)

Mohammad Z. Meybodi
Indiana University Kokomo
mmeybodi@iuk.edu

ABSTRACT

The focus of this article is to examine if lean manufacturing practices have positive influence on the alignment of strategic benchmarking performance metrics and organizational competitiveness. A set of nine hypotheses was used to examine the differences between conventional and lean manufacturing companies in setting organizational goals and objectives, scanning environmental factors, building core competencies, setting competitive priorities, establishing manufacturing objectives, and alignment of competitive capabilities with competitive priorities. Statistical results indicate compared with conventional companies, lean manufacturing organizations are overall better in recognizing external environmental factors, building organizational core competencies, focusing on sustainability, and aligning competitive capabilities with competitive priorities.

KEY WORDS: Benchmarking, Performance Metrics, Lean Manufacturing

INTRODUCTION

Benchmarking is an effective improvement tool utilized by organizations for improving aspects of organizational competitive priorities such as cost, quality, delivery, flexibility, and customer service. Benchmarking may be defined as a process in which an organization tries to learn from the best-in-class organizations, determine how the best-in-class achieve superior performance levels, and utilize those practices as benchmarks to their own organization ((Sahoo (2021), Yazan et.al (2019), Watson (1993)). The primary reasons for the use of benchmarking by world-class organizations may be summarized as: 1) Benchmarking is a flexible tool that can be used for gradual continuous improvement as well as for major changes of process reengineering (Bogan and English, 1994). 2) It is a valuable educational tool that provides opportunity to learn and prepare a company for change because it exposes employees to new approaches, systems, and procedures of other organizations ((Singh et. al. (2018), and Zairi (1994)). 3) It is an efficient tool to capitalize on proven ideas and to avoid the cost of additional resources for developing new ideas from scratch. 4) It is an effective tool for improving quality, increasing customer satisfaction while minimizing both the cost of good quality as well as the cost of poor quality (Blanchard, 2008).

Deming (1982) and a number of other quality advocates have strongly recommended the use of benchmarking as an essential component of continuous improvement ((Ahmed et. al (2023), Graham (1993), Ishikawa (1985), Venetucci (1992), Dawkins, Feeny, and Harris (2007)). Since 1987, benchmarking has been a major component of the Malcolm Baldrige National Quality Award criteria; it has consistently influenced more than half of the total Baldrige points ((Bogan

and English (1994), Ford and Evans (2001)). The practice of benchmarking is also being widely used for six sigma process and for organizations seeking ISO 9000 certification.
Lean manufacturing (LM) has been a great force in the world of manufacturing since the early 1980's. Some of the main benefits of lean manufacturing such as inventory reduction, quick delivery, quality improvement, and cost reduction have been well documented ((Cook and Rogowski (1996), Hobbs (1994), Temponi and Pandya (1995)). In the simplest form, LM requires production of the right parts in the right quantities and at the right times. Elimination of waste and respectful treatment of people are the two fundamental principles of a LM system. The focus of this article is to take advantage of the principles of LM and examine if there are relationship between LM practices and alignment of benchmarking performance measures.

LITERATURE REVIEW

Since the early 1980's a large number of articles have been written on the development and application of benchmarking in diverse areas of businesses such as manufacturing, health care, marketing, supply chain, energy, and customer service. Harrison (1999) presents an analysis of the evolution of different aspects of benchmarking activities. A comparison of the Xerox and Kodak benchmarking processes has been reported by Bogan and English (1994). Although the two benchmarking methods utilize a different number of steps, but their overall benchmarking activities are quite similar. Successful application of benchmarking at British Royal mail has been reported by Zairi and Whymark (2000). Applications of benchmarking to public procurement, world-class purchasing, and to the US service sectors have been reported respectively by Raymond (2008), Newman, Hanna, and Duffett (1995), and Roth et al. (1997). Bartley, Gomibuchi, and Mann (2007) utilized benchmarking to provide insights into how organizations can develop more customer-focused culture. Seong-Jong et. al. (2009) used benchmarking to measure the performance of a number of specialty coffee stores. Singh, Narain, and Yadav (2006) utilized benchmarking and performance measurement to investigate supply chain management practices at a number Indian manufacturing organizations. They found that Indian organizations were using benchmarking mainly as a continuous improvement tool. Chia et. al (2009) also employed a balanced scorecard approach to measure the performance of a number of entities in the supply chain. The authors concluded despite the need to utilize a balanced performance metrics; organizations primarily focused on the use of traditional financial measures. Gurumurthy and Kodali (2009) utilized benchmarking to assess the implementation of lean manufacturing. Practical application of lead benchmarking and performance measurement to achieve organizational change has been investigated by Moffett, Anderson-Gillespie, and McAdam (2008). Goncharuk (2008) investigated the capability of using performance benchmarking tools for estimation of efficiency in gas distribution companies. The use of benchmarking to measure operational performance of organization utilizing internet based services has been reported by Hadaya (2009). A comprehensive list of legal and ethical issues of benchmarking is presented by Brue (2002) and Vaziri (1992). The use of benchmarking as an effective organizational learning tool has been presented by (Hichem et.al. (2022), Malik et.al. (2021), Tasopoulou and Tsiotras (2017), Deshpande et. al (2023), Feibert et. al (2019), Souliotis et.al. (2017), Senge (1990), Ford and Evans (2001), Smith (1997), Hambly (1997), Watson (2001), Chen and Paetsch (1995), Gleich et. al. (2008), O'Dell and Grayson (2000), Routroy and Pradhan (2013), and Evans and Dean (2003)).

Although for the past three decades, there has been considerable volume of research on the application of benchmarking in various areas of business, but the primary focus of the research have been on short term technical and financial aspects of benchmarking. Fundamental strategic factors such as changing organizational culture, recognizing external environmental factors, and building organizational core competencies were generally disregarded. As argued by Goldwasser (1995), Kaplan and Norton (1992), Ahmed and Rafiq (1998) and Talluri and Vazacopoulos (1998) benchmarking is more than a comparative analysis of technical and financial metrics from one company to another. To be effective, organizations need to integrate benchmarking activities into long term organizational strategy and benchmarking process needs to employ a broad range of balanced performance measures that are consistent with organizational strategy ((Shukla, S.K. and Sushil (2022), Kaplan and Norton (2001)). The balanced performance measure approach provides a framework that helps managers to identify what mix of performance metrics needs to be measured that are aligned with overall organization strategy. In other words, the balanced performance measures enable organizations to clarify their mission and strategy and translate them into action. It provides feedback on multiple financial and non-financial performance metrics and allows organizations to consistently implement their strategy. In a benchmarking study article, Meybodi (2005) demonstrated the inconsistency of traditional organizations in choosing benchmarking performance measures at various levels of organization. That is, performance measures chosen by managers at operational levels were inconsistent with overall organizational strategy. The lack of a consistent strategy is a major impediment in building core competencies to ensure long term organizational competitiveness. The objective of this article is to examine if there are relationship between LM practices and utilization of a balanced benchmarking performance measures that are consistent with organizational strategy.

STRTEGIC BENCHMARKING PERFORMANCE MEASURES

Utilizing LM principles is an effective method to stay competitive in the market. Lean manufacturing focuses on reduction or elimination of all kinds of wastes such as lower inventory, faster production cycle time, more teamwork, and higher labor productivity. In a number of studies, researchers showed that organizations with effective LM system are successful not only in areas such as inventory reduction and quick delivery, but also in other areas such as quality improvement and new product development (Petrillo et. al (2019), Moori et. al (2019), Handfield (1993), Pettersen (2009)). As argued by Cook and Rogowski (1996), to understand application of LM in other areas, one has to carefully examine its two principles of eliminating wastes and respecting people. Looking at LM as a process for eliminating wastes and respectful treatment of people; its principles can be applied to other areas such as benchmarking performance measurement. Managers across various industries practicing total quality management and six sigma have recognized that effective performance measurement is the key for organizational success. Without effective performance measurement a company doesn’t know the real problem, who is responsible, where improvement efforts are needed, and the amount and type of necessary resources. The special focus of well-known quality improvement and quality award processes such as six sigma, ISO 9000, Malcolm Baldrige National Quality Award, and Deming prize on benchmarking and performance measurement is perhaps a clear indication of the critical role of these elements in managing and improving organizational processes. As argued by Eccles and Nohria (1992), in the past organizations used performance measurements that contributed primarily to short-term financial and technical results. How the organization achieved those results and their impact on the entire organization was secondary. Today, managers understand that focus on short-term financial and technical results without consideration to overall organizational strategy could produce devastating results over the long term. As a result, organizations are learning to manage the system in a way that crosses departmental boundaries. In this new horizontally integrated system, organizations need to accept a long-term perspective and to utilize balanced, broad reaching financial and nonfinancial performance measures to carefully improve the competitiveness of the entire organization. This approach requires that benchmarking organizations develop a complete understanding of their own business strategy and deployment of the strategy into functional strategies. This process will ensure that there is a consensus within the organization about long term and short-term performance measures that are consistent with organizational mission and goals (Lockamy (2019), Papke-Shields et. al. (2002), Hoque and Maalouf (2022), .and Madigan (1992)). By using measurable performance results with focus on the entire organization, managers will be able to determine their progress toward long term goals and objectives. With respect to the preceding discussion on performance measurement, the article presents the following hypotheses:

H1: Lean manufacturing companies are more consistent than conventional companies in setting their balanced long term goals and objectives?

H2: Lean manufacturing companies are more consistent than conventional companies in scanning environmental factors to set their goals and objectives.

H3: Lean manufacturing companies are more consistent than conventional companies in building their core competencies to support their goals and objectives.

H4: Lean manufacturing companies are more consistent than conventional companies in aligning their competitive priorities with their goals and objectives.

H5: Lean manufacturing companies are more consistent than conventional companies in placing a higher emphasis on elements of flexibility, agility, and time based competition as their competitive priorities.

H6: Lean manufacturing companies are more consistent than conventional companies in aligning their competitive capabilities with their competitive priorities.

H7: Lean manufacturing companies are more consistent than conventional companies in aligning their manufacturing objectives with their competitive priorities.

H8: Lean manufacturing companies are more consistent than conventional companies in placing a higher emphasis on fundamental strategic manufacturing objectives.

H9: Lean manufacturing companies are more consistent than conventional companies in placing a higher emphasis on environmental and sustainability issues.

RESEARCH METHODOLOGY

A questionnaire-based mailed survey was used to test the above hypotheses. The survey contained a series of questions on the use of strategic and operational benchmarking performance measures for conventional and LM organizations. Strategic questions are concerned about organizational mission and goals, as well as attitude toward customers, competition, technology, globalization, development of core competencies, and organizational competitive priorities. Operational items are related to specific technical performance measures such cost, quality, and delivery.
The target population for this study consisted of manufacturing firms in Midwestern United States. A sample of 500 manufacturing firms was chosen from manufacturers' directories of the midwestern states. In addition to general organization and managerial profile items, the survey contained series of questions regarding organizational goals and objectives, competitive priorities, competitive capabilities, manufacturing performance objectives, manufacturing action plans, and LM practices. Out of 91 completed surveys received, 84 surveys were usable

resulting in a response rate of 17%. Out of 84 usable surveys, 33 respondents declared their organization to be a LM organization and 51 respondents considered their organization to be a traditional organization.
The survey data indicates the majority of respondents had various levels of managerial positions of organization with less than 500 employees. Presidents and vice presidents accounted for 29% and plant managers accounted for 30% of the sample. About 35% of the sample had other managerial positions such as operations/production managers, quality managers, and the remaining 6% were production line supervisors. In terms of manufacturing experience, about 28% of the respondents had between 10 to 20 years and 60% had more than 20 years of manufacturing experience.

RESULTS

Table 1 and 2 show respectively the mean importance ratings for corporate goals and objectives and strategic environmental and core competencies factors. The respondents were asked to rate each factor based on the degree of importance (1=low importance, 5=high importance) to the company for the next five years. Table 1 indicates that, for conventional companies, the ranking for the corporate goals and objectives are building market share, maximizing profits, focusing on customer satisfaction, and building sustainable organization. Focus on customer satisfaction, building sustainable organization, building market share, and maximizing profits are the corresponding rankings for LM companies. Being in a sustainable competitive position with respect to customer satisfaction is a possible explanation for market expansion and profit making posture. Statistical result of Table 1 shows although the ranking of corporate goals and objectives for conventional and LM companies are not quite the same but as the last column indicates with the exception of building sustainable organization none of the three tests are significant; hence, overall, the first hypothesis is not supported by the data.
Table 2 shows for conventional companies, the ranking of the strategic environmental and core competencies factors are focusing on competition, building innovative and agile organization, developing strategies for sustainability, knowledge workforce, global issues, and technology. However, the ratings for these factors are not as strong as the ratings of the corporate goals and objectives in Table 1. This is perhaps an indication of a typical reactive strategy by conventional companies in which the primary focus of managers is on customer satisfaction to increase market share and profits. Understanding external environmental factors such as competitors' strategy, sustainability, global issues and building innovative and agile organization through development of knowledge workforce and state of the art technology to effectively deal with environmental factors are not under prime consideration. For conventional companies, understanding the causes for such strategic misalignment between corporate goals and objectives and detection of external factors as well as proactive development of organizational strategic core competencies is crucial. Table 2 shows for LM companies, building innovative and agile organization, developing sustainable strategies, strategic workforce as the top three strategic environmental factors. These are closely followed by understanding of competition, technology, and global issues. The last column of Table 2 clearly shows that for LM organizations, the mean rating for these factors are significantly higher than the mean ratings of the corresponding items for conventional organizations. This is an indication that, unlike conventional companies, LM organizations focus more on understanding external environmental factors and especially building organizational core competencies through development of knowledge workforce and state of the art technology. In fact, LM organizations often develop their core competencies first and then utilize a proactive strategy to find opportunities for exploiting their core competencies to develop new products and services to achieve competitive advantage in the market. Statistical result of Table 2 clearly supports hypotheses H2 and H3.

The rating of competitive priority factors for conventional and LM organizations is shown in Table 3. From Table 3, the respondents from conventional companies ranked product reliability, conformance quality, delivery reliability, price, and fast delivery as the top five important competitive priorities. The ranking of product reliability and conformance quality as the top two competitive priorities is consistent with corporate strategy. It indicates the responding managers believe that quality factors are still important factors of competition. However, the ranking of delivery reliability, price, and fast delivery as the next three competitive priorities indicate that managers also believe on the importance of delivery and price. Relative low ranking of factors such as customization, new product development speed and sustainability is inconsistent with the corporate strategy of customer satisfaction through building innovative, agile, and sustainable organization.

the right side of Table 3 shows that the respondents from LM companies ranked product reliability, delivery reliability, fast delivery, sustainability, new product development (NPD) speed, and product customization as their top six important competitive priorities. The ranking of product reliability as the top competitive priority indicates that managers of LM companies also believe on the importance of quality as an essential factor of competitive advantage. However, the ranking of delivery reliability, fast delivery, sustainability, NPD speed, and product customization as the next five competitive priorities indicate that the respondents also believe on the importance of time based competition, agility, environment, and product customization. Table 3 also shows that conformance quality and price as factors of competitive priorities ranked relatively low by LM companies. This rather interesting result indicates that, unlike conventional companies, the responding managers from LM companies believe that conformance quality and low price are no longer the primary competitive advantage factors. From the results of Table 3, one can conclude that overall LM organizations are more consistent than conventional organizations in aligning their competitive priorities with their corporate goals and objectives. Also, from table 3 it is clear that LM companies place more emphasis on the factors of time based competition, agility, sustainability, and product customization. From statistical results of Table 3, we may conclude that the data supports hypotheses H4 and H5.

To help understand relative strength of organizational competitive capabilities, the respondents were asked to rate relative competitive strength of their organization with respect to the competitors who are doing best for each element of competitive priorities. A five-point scale, where 1 corresponds to weak and 5 to strong, is used to indicate managers' perceptions of the company's current competitive capability relative to the best competitors. The mean strength scores for each element of competitive priorities for conventional and LM organizations are shown respectively in Table 4 and 5. As the last column of Table 4 indicates, for conventional companies, the mean strength of the top five competitive priorities is significantly lower than the mean importance. This indicates that for these companies, although managers ranked product reliability, conformance quality, delivery reliability, price, and fast delivery as the top five competitive priorities, organizational capabilities of those elements is not that strong. This imbalance between organizational competitive priorities and their competitive capabilities is a serious area that needs to be investigated. Statistical results of Table 5 indicate the mean strength of the elements of competitive priorities for LM companies is significantly higher than conventional companies. In fact, statistical tests indicate that, unlike conventional companies, the mean strength for majority of the elements of competitive priorities is close or higher than the mean importance ratings. This indicates for LM companies there is a better balance between competitive priorities and organizational competitive capabilities and hence the data supports hypothesis H6.

Table 6 shows the ratings of manufacturing objective factors for conventional and LM companies. Table 6 suggests for conventional companies the future role of manufacturing unit is to focus on improving conformance quality, product reliability, and supplier quality. Reducing the costs of labor, overhead, materials, set-up, and increasing capacity utilization are the next set of objectives. The first three objectives are consistent with competitive priorities and the strategic importance organizations are placing on customer focus as a means to expand market share and maximize profits. However, the early emphasis on reducing costs and increasing capacity utilization for the next five objectives seems to be misaligned with the competitive priorities. Statistical results in the last column of Table 6 supports this statement by demonstrating that the mean importance ratings of these cost factors for conventional companies is significantly higher than the corresponding cost factors for LM companies. Such inconsistency may be due to miscommunication among managers at various levels or the results of inconsistent evaluations and reward systems. That is, while organizational strategy calls for customer focus, quality, agility, sustainability, and customization, production managers are often rewarded based on efficiency measures such as cost cutting or capacity utilization. Table 6 also indicates relatively low emphasis on strategic factors such as changing organizational culture, improving inter-functional communication, improving employee empowerment and morale, and improving supplier relationships. Since these are the foundation of world-class organizations, there needs to be higher emphasis on these factors. Also, there should have been a higher emphasis on important competitiveness factors such as eliminating waste, improving NPD speed, improving environmental issues, and reducing inventories.

Table 6 shows, unlike conventional companies, LM organizations place higher emphasis on fundamental organizational factors such as changing organizational culture, improving inter-functional communication, improving employee morale, improving teamwork, and improving supplier relationships. This is followed by focusing on competitiveness factors such as eliminating wastes, improving quality, improving delivery and customization, reducing inventory, and increasing capacity utilization. The ranking of these factors is consistent with the principles of LM and LM research. Statistical result in the last column of Table 6 shows the mean importance ratings of these fundamental factors for LM companies is significantly higher than the corresponding factors for conventional companies. Statistical result of Table 6 clearly supports hypotheses H7 and H8.

For hypothesis H9, we needs to reexamine preceding results from Tables 1-6. The results from these tables showed that while the focus of conventional companies was on a number of quantitative cost reduction and quality improvement factors, LM organizations utilized more balanced quantitative and strategic qualitative factors. Specifically, LM organizations placed more emphasis on sustainability and environmental issues; and hence the data clearly support hypothesis H9.

CONCLUSION

The focus of this article was to examine if there are differences between conventional and LM organizations in the alignment of their strategic benchmarking performance metrics. Summary of statistical results indicate the following:

- LM organizations are better in recognizing external environmental factors such as competition and global issues to set their strategy.

- LM organizations are better in building their core competencies to effectively deal with the external environmental factors.

- LM companies are more consistent than conventional companies in aligning their competitive priorities with their goals and objectives.

- LM companies are more consistent than conventional companies in placing a higher emphasis on strategic factors of flexibility, agility, and time based competition.

- LM companies are more consistent than conventional companies in aligning their competitive capabilities with their competitive priorities.

- LM companies are more consistent than conventional companies in aligning their manufacturing objectives with their competitive priorities.

- LM organizations are more consistent than conventional companies in placing a higher emphasis on fundamental strategic manufacturing objectives.

- LM companies are more consistent than conventional companies in placing a higher emphasis on environmental and sustainability issues.

Managerial implication of the article is that successful implementation of LM principles goes much beyond traditional manufacturing benefits of inventory reduction and frequent deliveries. For LM organizations success in their strategic benchmarking performance metrics is the result of knowledge and technology transfer from their LM system into their organizational strategic plan.

(Tables and references are available from the author upon request)

DECISION SCIENCES INSTITUTE

Leveraging Blockchain in the Supply Chain to Compete as a Small to Medium Enterprise Manufacturer

Brian Laird, *Towson University*
Email: BrianSLaird88@gmail.com

Barin Nag, *Towson University*
Email: bnag@towson.edu

ABSTRACT

Small and Medium Enterprises (SMEs) often face difficulties in adopting new technology like Blockchain that is changing business in the supply chain. In this study, a Lithium Ion Battery (LIB) manufacturing SME was analyzed to study the typical problems of data and inventory management to identify potential Blockchain applications. The study is an extension of previous in general studies on SMEs and looks specifically at inventory management inefficiencies. The quantitative analysis of data gaps in material procurement and tracing leads to benefits from the enhanced transparency and tracking capabilities of Blockchain for greater efficiencies and competitive advantage.

KEYWORDS: Blockchain, Lithium-Ion Battery, SME, Material Procurement, Inventory Data, Supply Chain Management

INTRODUCTION

Supply Chain Management (SCM) has risen as a major business discipline in the last few decades alongside an emphasis on data analytics. SCM typically focuses on managing the flow of goods, services, and information in and out of an organization with a goal of lowering total costs at a given minimum acceptable service level. Forecasts driven by accurate and reliable data are critical to inventory policy. The inventory strategy taken by a company must align and integrate into the overall strategy of the business (Nag et al.,2014). Too much inventory is inefficient and expensive, too little can run the risk of a stockout and missed sales. Supply chains extend beyond individual firms, they create value chains from raw materials processors down to the customer/consumer level. They have attributes such as speed and resilience, which is their ability and tolerance to handle interruption while fulfilling their function (Shao & Jin, 2020). In modern economies, firms no longer compete with one another; their supply chains do.

Large corporations are not the only members of global supply chains. Companies designated as Small and Medium Enterprises (SMEs) face the same challenges of globalization and changes in the technological landscape as large corporations do. There is no universal definition, but SMEs are typically classified as independent companies that employ fewer than 250 people. SMEs form the majority of businesses in many countries and are especially important in the economies of developing nations. They usually have less tolerance for risk and must operate very efficiently to survive alongside larger competition (Ethem & Susanne, 2019). SMEs must compete for the same materials and customers as their competitors do, only with less resources to do so.

As certain industries are poised for tremendous growth in the coming decades, SMEs must adopt the technologies of Industry 4.0, as the new industry paradigm is called, to remain relevant in the marketplace and scale with demand. This study focused on one emerging technology, Blockchain, and how it can be used to address common problems faced by SMEs in the manufacturing sector. A literature review was performed on Blockchain technology, SMEs, the Lithium-Ion Battery industry, and current integration methods for Industry 4.0 technology. Additionally, raw material data was collected from an SME lithium-ion battery manufacturer and analyzed to demonstrate two typical SME problems. Two specific raw materials were isolated for further analysis by gathering information from the company Enterprise Resource Planning (ERP) system, their internal network, and from supplier partners. By analyzing this case data, suggestions for Blockchain technology-enabled solutions were provided and the potential benefits explained. This exercise can be scaled up or replicated as a template for other SMEs to discover their own critical areas to target for Blockchain mediation. This study was not a technical dive into how Blockchain works, but rather a proposal of the many benefits that can be derived from proper application to specific SME problems.

PROBLEM STATEMENT

Globalization has been greatly accelerated by the internet and has increased the complexity and competition levels within today's supply chains. In manufacturing, organizations from around the world are now enabled to compete in the same supply markets for the same resources, including finite natural resources. Manufacturing industry supply chains have evolved from linear, single node supply chains with clear up and down-stream paths to a community web exchange of goods, services, and information (Jiang et al., 2022). Global competition makes profit margins shrink and efficiency becomes a priority. As has always been the case, organizations that fail to evolve with the times are doomed.

SMEs typically face significant barriers that prevent them from acting like their larger competition. They usually run lean, efficient operations and position themselves to adapt and innovate. Despite their best efforts to evolve with technology, many are plagued by problems that come with antiquated supply chain practices. Processes may be heavily paper-based or email-based. Data entry may be decentralized and subject to input errors, information gaps, redundancy, and non-consensus between multiple ERP systems. The number of parties involved in handling data opens the door for fraudulent behavior and cultures of dis-trust between parties. Low trust policies with logistical providers and vendors slows down the flow of goods. An excessive amount of on-hand inventory may be carried to mitigate other risks. These inefficiencies become costly to the company and a typical SME may lack the resources to address them.

Blockchain technology is already being integrated into modern supply chains and is accessible to companies of all sizes. There is a need to find ways for SME manufacturers to use Blockchain to improve efficiency in securing raw materials and succeeding among large, global competitors. In this context, this study looked at the supply chain efficiency problems SMEs are experiencing and how Blockchain solutions can be implemented to alleviate them. These problems are especially important in environments where demand is growing, where production needs to be scaled up, and competition is increasing in raw materials procurement. Systemic problems often compound each other. Addressing one will often alleviate stress in other areas and reduce total operational costs.

Quantitative data from an SME categorized Lithium-Ion Battery manufacturer was gathered and analyzed to identify the prevalence of inventory practice inefficiencies amongst critical raw materials that are purchased and used in production. Two materials were systematically identified for their significant importance in the anticipated growth from rising industry demand for the end products. Data was collected on these materials to map out blind spots in the upstream visibility of the material information. A qualitative analysis was conducted to determine why these information gaps may exist.

All analysis was done with a focus on upstream supply to demonstrate two common problems experienced by SMEs, i.e., insufficient data collection and inefficient inventory practices. Thus, further attention was drawn to the most critical materials whose risks of supply availability pose a significant threat to operational continuity. Supply chain risks include any threat to the flow of information, raw material, and finished goods throughout the supply chain. They include events or conditions at any level that may lead to operational, tactical, or strategic failures that disrupt business performance (Lai et al., 2021). Also included are threats to the health and environment of participants. By observing the existence of these typical SME problems in the context of a real business case, conclusions were drawn as to which and where Blockchain solutions were most needed and could be the most effective. Recommendations were made to facilitate Blockchain adoptions to help the SME compete in a growing and competitive marketplace. These recommendations can also broadly apply to most other SMEs as they use Blockchain to address their own organizational inefficiencies.

BACKGROUND & LITERATURE REVIEW

Blockchain Defined

Blockchain is a digitized version of a distributed ledger, comparable to a standard ledger of accounting practice, that operates with decentralized control. It is built upon programmable code to run across all linked computer networks. Instead of central authority holding and verifying the records, a decentralized peer-to-peer digital network maintains copies of the ledger across all the nodes on the network. The nodes are made up of authorized people or systems that were granted access to the network via an encrypted key or other identity verification. The decentralized network keeps a ledger of a continuous chain of blocks of data, or hashes, that keep a chronological, time-stamped history of all changes made to the data on the chain (Jiang et al., 2022). No single entity controls the nodes or can modify rules. The effect is a complete audit trail of all transactions that is visible to all authorized users of the network (Madhani, 2021).

Blocks of data cannot be permanently modified by a single party without the full consensus of the change by the entire network of ledgers. This level of encryption makes Blockchain extremely secure from fraudulent manipulation. Since data is not stored in one centralized server, the decentralized nature of the data makes it less susceptible to attacks. This is the technology that popular crypto currencies like BitCoin are built upon, but Blockchain's uses extend far beyond currency and financial record keeping. Blockchain can be used to keep a verified, immutable history of any piece of data shared between multiple parties. The more parties that have access to the data and keep a ledger, the stronger the chain is. Blockchain can be used to verify shipping manifests, confirm the chain of custody of a piece of artwork, or certify the recycled content percentage of a product, just to name a few applications.

Blockchain Qualities

The most notable qualities and characteristics of Blockchain technology are transparency, immutability, verifiability, traceability, decentralization, and security. Each was discussed below. Transparency of data granted by Blockchain allows all users to see all information stored on the ledger. This means data can be accessed from top to bottom by users on a supply chain, such as manufacturers or end customers. Users can learn from this data and use it as they wish, such as to get tracking information on a shipment. Data is accessible and not hidden within siloed systems. It is available without loss, noise, delay, or distortion (Sunny & Madhusudanan, 2020). Useful information about activity of other members of the network is not hidden or made unavailable.

The Immutable, permanent nature of data stored in hashes is one of Blockchain's foundations. All transactions are permanently recorded and cannot be changed without group consensus, making the data essentially tamper-proof (Xu et al., 2021). The authenticity of the data empowers decision making with higher confidence levels.
By nature, Blockchain provides multiple points of confirmation on a piece of data. Essentially, any data on the chain can be verified by multiple parties as being accurate. This provides consensus without the need for a third party to provide the validation of transactions or any transfers of value (Hashimy et al., 2021). Any instances of non-consensus are investigated, and the unverifiable claim is rejected. This allows any party in the network to make a highly credible claim that can quickly be verified as true.

Due to immutable historical record keeping, data stored on a Blockchain ledger is traceable back through time. All previous versions of the ledger are permanently kept on record and cannot be destroyed or modified without full consensus. Problems can be researched to their source and physical goods can be traced throughout their lifecycle (Wu & Zhang, 2022). Any linked item, currency, or other piece of data can be tracked through all changes of custody between parties on the network. Materials can be traced back to their source in the ground which enables enhanced carbon tracing and responsible sourcing practices.

The decentralization of the ledger is one of the key qualities that makes this technology so superior to traditional ledgers. This could not be accomplished without the internet and the current digital age of data access and storage. Data is entered via a decentralized node and distributed to the entire network for confirmation. There is no centralized power of authority over the data, but certainty in the data is still established as it would have been in a traditional arrangement (Jiang et al., 2022). All data generated by the network is packaged into blocks and copied to be stored in a distributed fashion (Wu & Zhang, 2022).

As a data management technology, Blockchain needs to be secure in order for any party to consent to using it. Blockchain uses cryptography and programmed verification to ensure any party that accesses the network with their private key is the proper proprietor of the information (Cordova et al., 2021). Although data is shared with all nodes, private Blockchains can be used to limit the scope of what is fully disclosed at any given time (Antônio et al., 2022). The cryptographic nature of the technology is what ensures immutability of the linked blocks. A number of security tools used by the owners of the information can be enabled and integrated within the technology such as hash functions and private key signatures (Jiang et al., 2022). Third party security custodians can be included on the chain network. Many of these defense tools and the distributed nature of the data provide high levels of protection from nefarious actors who attempt to access or alter network data.

Blockchain Business Benefits

A business who chooses to integrate Blockchain technology into their operations will derive countless benefits, with many more that will be developed in the future. Some of the main benefits were discussed in this section.

Blockchain has proven to be a major disruptor in finance practices. Participating organizations and financial institutions can link their accounts payable and receivable on Blockchain platforms to execute and verify transactions when the proper conditions are met. Digitized invoices can be verified, and electronic transfer payments can be executed automatically. This reduces traditional transaction costs. It can also eliminate the need for banks, clearing houses, and forms of third-party credit verification (Jiang et al., 2022). Digital, “crypto” currency itself can also be used in the transactions.

Ledger verified data can be shared with creditors for better credit ratings and the ability to access capital. Banks can verify accounts receivable on a real-time, continuous basis and obtain better loan interest rates to cover accounts payable. Creditors will have less manual review to conduct, giving the applicant better odds at even being eligible for obtaining credit. Credit can also radiate to other members of a supply chain based on verified contracts and endorsement of debts from partners (Xu et al., 2021). This trust transmission, also referred to as credit penetration, helps members of a supply chain endorse each other through network generated credit. By understanding a company’s relationships and place in a supply chain, they may be seen as less of a risk by creditors (Jiang et al., 2022). Seeing other suppliers’ credit can help a firm select a more cooperative partner with less delinquency on payments. Improved supply chain network visibility influences partner financial behavior. Bad behavior is visible and broadcasted through the network which incentivizes parties to improve their financial practices if they wish to remain in the network (Zheng et al., 2022). This, in turn, provides all parties a cost savings by not having to find new partners.

Operationally, Blockchain provides many ways for a company to be more efficient. The presence of more high-quality data that is shared and compiled in the network allows for higher process coordination and linked forecasts among partners. Forecasts can be built on real-time location data of materials and processes that can be traced and verified. Combined with other automated tools, resources throughout a company and throughout the network can be more efficiently allocated. More participants are incentivized to use the Blockchain network making it stronger. With less need for data entry, redundant job roles and redundant data centers can be removed from processes, organizations, and the supply chain (Kumar et al., 2021). In coordinated processes, Blockchain allows for a clear delegation of responsibility, and a relief for non-responsible parties. Interoperability allows for supply disruptions to be handled efficiently through the whole chain. Within an organization, it also removes data silos and links common data together through mutual confirmation. Multi-point verification of the steps in order fulfillment can lower the chances of mistakes and speed up the entire procure-to-pay cycle. This cycle had to previously pause and wait for manual permission to proceed between steps (Lai et al., 2021). It is estimated that at least 33% of all supply chain activities can be improved by Blockchain technology (Madhani, 2021). Data in the system is enabled to stay in constant feedback loops as information and material changes hands, which allows for the opportunity of continuous improvement, weakness identification, and stronger data-driven decision making. Data errors are reduced, and real-time data updates are further enabled, thus providing a more ideal data source for machine learning tools.

All companies that do business together have to have trust in each other at some level. Having low trust is very costly; companies need to proceed with excessive caution which slows down all processes. High levels of trust and cooperation keep operations running smoothly, but there is risk involved. Blockchain technology is an effective trust enabler, enhancer, and facilitator tool. Here, trust refers to the expectation that network nodes will continue to behave with similar ability, honesty, and reliability based on their historical transaction records (Wu & Zhang, 2022). A great benefit of Blockchain is that businesses can work together with low to almost zero actual trust in the traditional sense. Businesses that are completely unknown to each other can rely on Blockchain to regulate their relationship, especially when it comes to financial exchanges. Intellectual Property can remain safeguarded since the original owner can use Blockchain data to prove ownership in court. This all also provides external investor trust and transparency. Additionally, organizations can better authenticate and guarantee their products and business practices to their consumers. They can provide an immutable product history with extensive detail so consumers can trust what they are buying without the need of a third-party verification.

The level of verified data that Blockchain collects and maintains is of great use to any type of regulatory body. The data will also be useful in future regulatory compliance as laws change, such as carbon emissions requirements for processes and products (Antônio et al., 2022). The data is also highly auditable for tax purposes. It is even possible that a company can become eligible for government grants for tracking their data in this fashion.

Blockchain allows for better collaboration through enhanced data sharing while working with other businesses on projects. Problems can be resolved quickly by investigating discrepancies in the shared ledger. This is a great aid in conflict resolution and can be done so without costly legal processes. Due diligence in partner selection is improved due to high quality, reliable supplier data. There are also higher safeguards to prevent fraudulent behavior. Counterfeit products have more difficulty entering the supply chain since products and materials can be tracked and verified (Deberdt & Billon, 2021). Participation in a Blockchain system creates incentives for good behavior and prevents bad actions. Having Blockchain enabled processes in place can itself be a competitive advantage for a company in the marketplace.

Blockchain Business Tools

Blockchain is the underlying technology powering many novel business tools. First, there are digital currencies and tokens. Cryptocurrencies can be used as a medium of exchange with no centralized management or intermediaries. This creates added value and efficiencies in finance. Modern businesses can accept and issue digital payment. Debts and contracts can be made into digital tokens and traded like securities and futures options are traded on traditional public exchanges (Xu et al., 2021).

Marketplaces used to qualify and select suppliers have been built and improved by using Blockchain. The data curated by Blockchains have created entire information ecosystems of auditable data about supplier offerings and performance history, thus automating formerly labor-intensive process. This live market can showcase verified statistics on vendors so that a buyer can more efficiently qualify them before doing business, lowering the barriers to onboarding and the costs of building relationships. Buyers and sellers of products and services can understand the market and switch partners with lower costs of doing so. Additionally, multiple buyers can pool resources to run in depth audits on supply chain partners instead of each party funding their own research. Instead of only having static information, audits can be done on a continuous basis and companies can constantly monitor the suppliers they use, something that is typically far too expensive for one company to do on their own (Hildebrand & Schneider, 2020). In every industry, to varying degrees, there have always been incentives for bad actors to insert themselves into a supply chain. With a continuously audited supplier network, bad actors can be ousted faster or discouraged from participating to begin with. Other tools can also be used among supply chain parties such as reputation management functions. This public review system can be used to grade suppliers and give them a public rating (Xu et al., 2021).

Digital Custody Chains are a Blockchain enabled tool that has improved the tracking of physical goods through complex systems. Since information can be distributed and verified in real time using Blockchain, a whole host of “hyperconnected manufacturing” tools can be used today. Continuous supply of reliable data allows equipment and production automation decisions to be made on the fly in response to changing forecasts (Morteza & Mohammad, 2021). Trusted data can accelerate through supply chain systems from the customer to the manufacturing floor. Other Industry 4.0 tools, namely Internet of Things (IoT) and Radio Frequency Identification (RFID), are used in conjunction with Blockchain technologies. IoT is the integration of machines and other tools with the internet to form a connected network. For example, sensors and readers in the bed of a tractor trailer can feed location and temperature data of cargo as it moves across the country. RFID tags are applied to physical goods so their location and status can be updated regularly at scale for advanced inventory control. Hyperconnected manufacturing integrates manufacturing with horizontal and vertical partners of the supply chain.

Blockchain enables a new type of “smart contract” between parties that allows for more efficient fulfillment. Contracts can be digitally programmed between parties to execute certain actions automatically when proper conditions are met. Blockchain will verify these conditions through a consensus. This can reduce the need for third party verification systems like banks or legal entities. This lowers transaction costs and allows for easy contracts between parties who may have low trust for one another (Ethem & Susanne, 2019). In a game theory simulation study conducted by Zheng et al. (2022), researchers set parameters of a smart contract between automated nodes of a Blockchain. The nodes were programmed to act in their own self-interest to seek highest utility and lowest operational costs in a simulated supply chain. Each node was given the ability to either follow the agreement and perform their contractual function or to default on their duty in the contract. A mechanism for reward and punishment was designed in the Blockchain in line with the smart contract. The study found that rationally acting nodes chose to perform the contract, none acted in self-interest by defaulting on their duty since there was far less incentive to do so.

Blockchain Integration

Blockchain as a technology tool is not a substitute for an Enterprise Resource Planning (ERP) system, but rather a plug-in or tool that can be used to support and supplement the data being managed. It has been argued though that a Blockchain system is superior to ERP for the production of highly customized products, since the goods and production processes need to constantly change (Espinoza et al., 2022). Given the vast array of ERPs that exist to run companies, Blockchain is most integrated using a middleware software that is developed and maintained by a third party rather than the user or ERP system itself. Third party services can also be used to handle Blockchain participation by proxy for a company (Sunny & Madhusudanan, 2020). Logistical support companies are also important to integrate into the network since they are an important supply chain service that many companies outsource.

Adoption of this technology by a company is complex and reliant on many factors. Despite this, it has been estimated that by 2025, 10% of the global economy as a whole will be using Blockchain technology in their business functions, generating a value of 3 trillion dollars (USD) (Kumar et al., 2021). Organizational adoption determinants include technological maturity, top management support, cost concerns, perceived usefulness, external support from vendors, change management competency, and resource availability (Morteza & Mohammad, 2021). An important consideration regarding Blockchain is that it only works when multiple parties are willing to participate in it; its strength is proportional to the number of users in the network.

Before initiating a Blockchain network, a company can check if any big players in the supply chain are already using one and see if they can be integrated in.

Blockchain Challenges

Blockchain is still in its relative infancy, which needs to be taken into consideration. The technology still has drawbacks and challenges and may not be the right solution for every company and industry. Transaction verification is not always as instant as the information availability. Nodes on internet locations such as China are stuck behind slow firewalls. The massive amount of constant data processing also has very high energy costs that are financially and environmentally taxing. They can also be expensive systems to implement and maintain with skilled labor; most systems are still complicated to use for every-day workers. Smaller businesses must weigh the technology benefits with the high fixed and variable costs to build and maintain the necessary infrastructure. They may elect to participate via a third-party proxy for some cost savings. Despite its strengths, security vulnerabilities still exist, and the systems are not completely impenetrable to bad actors. Some companies that advertise Blockchain solutions are also scams themselves (Liu et al., 2020).

A major hurdle for the technology becoming more mainstream is that certain companies refuse to participate in Blockchain programs because they don't want to overshare their safeguarded information. For example, they may not want their prices made public to all customers and competitors due to the problems this can create. This makes the technology harder to use because it is reliant on a minimum number of participants to function properly. Companies that may be ready to adopt Blockchain could be held back due to the technological limitations of their partner companies that they want to use it with. In this way, internal and external support is required for adoption. Once adopted, the technical rollout is very slow and return on investment may not be quickly realized (Lai et al., 2021). Legislation and regulation is also still lagging behind the technology and is likely to change (Mugurusi & Ahishakiye, 2022). A lack of standards and regulations regarding the handling of identification and intellectual property is still an adoption deterrent. The technology is struggling to get off the ground in less developed nations for some of the considerations provided above.

About SMEs

SMEs generally employ fewer than 250 people (or 500 per the European Union's designation) and do less than $50 million in annual gross revenue, to give a general perspective of their size (Ethem & Susanne, 2019). They are the backbone of most countries' economies, and their success can be vital to economic health. In China, SME sized businesses account for over 94% of the total number of enterprises and their activity accounts for 60% of the total GDP (Wang et al., 2020). They are the majority class of businesses in the world. While they may lack in power and resources, they retain certain strengths and advantages compared to larger businesses.

Being smaller sized operations, SMEs can act with agility in many ways. They typically have less formalized processes and "red tape" to clear in management bureaucracy to enact a project or other action. They can change course based on new direction much quicker than a large company's momentum and stakeholders allow them to do. In manufacturing, an SME may be able to change over its production to a new item or accommodate a change with a relative speed advantage compared to a large organization. An SME may also not be bound to structured demand forecasts that are slow to accommodate new information.

SMEs typically run naturally lean to survive amongst many competitors. They also have less processes to optimize, so they can focus on their competitive advantages (Ethem & Susanne, 2019). An SME might thrive in a niche market they cornered using a strong competitive advantage and expertise that bigger players have not bothered to learn.

Financially, SMEs usually have less investors to answer to, as opposed to something like a large publicly traded corporation. They also are eligible for more state aid, grants, and tax breaks than what can legally be provided to a larger company in most countries.

SME Limitations and Problems

The relative size and power of SMEs in global marketplaces burdens them with disadvantages and limitations that can manifest as operational problems. It is important to note that many SMEs are small because they are young and inexperienced. Their problems can be summed up as resource availability, financial challenges, supply chain struggles, and inefficiencies. It can be difficult to treat these problems as independent events since they all cause, influence, and magnify one another. They will be discussed in this section.

The resources that SMEs have issues accessing are human resources, materials, and financing. By many legal definitions, SMEs lack a large workforce as a resource. They typically have fewer people per job role that needs to be performed. Workforce job specialization may be lacking for that reason. As a result, many functions are outsourced to costly intermediaries. Tighter budgets may restrict payroll which affects talent acquisition and retention. Smaller budgets and the risk averse nature of SMEs may prevent them from investing in emerging technology or keeping up with technological advancements, aside from funding a proper IT department. Studies in organizational digital transformation success by Morteza & Mohammad (2021) found that SME manufacturers lag other industrial business categories when it comes to Industry 4.0 transformation. Less than 5% of Korean SME manufacturers surveyed classified as smart manufacturing adopters, and Industry 4.0 adoption in two other subject Asian countries was under 20%. Companies may be interested in technology but lack the human resources to do so.

Material resources are less available to SMEs for many reasons. They may lack the full logistical capabilities required to secure resources or the facilities required to store and maintain inventory. Their size as a small company likely limits their negotiating and purchase power on the market. This causes high prices and a dependency on fewer suppliers who will do business with them, especially if the SME buys in relatively small quantities. Dependency on fewer suppliers puts the company at the supplier's mercy. As a smaller buyer, SMEs are often treated as second class customers to larger business-to-business accounts.

Capital resources are limited to SMEs for reasons that tie into their financial issues. SMEs take in less revenue than bigger companies and therefore operate on smaller budgets. They typically do not issue stock or have many major investors. Their typical risk-averse nature and obstacles to obtain larger loans adds to the problem. Risk aversion strategies may also prevent an SME from doing business with a relatively unknown partner and lose the opportunity of a successful relationship (Ethem & Susanne, 2019).

A lack of capital resources can often prevent a company from properly managing their credit standing, especially if they fund their business through debt. Without a strong credit history, an SME may be subject to up-front deposit requirements for purchases or very poor net payment terms. This puts them at risk of receiving low level products or services without much recourse or leverage for corrective action. As a lone enterprise without much of a verifiable track record, lenders often view SMEs as an uncontrollable risk not worthy of a manual credit review (Wu & Zhang, 2022). An SME that is too informally linked to a large, well established "core enterprise" may not be able to benefit from the transitive trust that is radiated out in a supply chain in the eyes of creditors (Jiang et al., 2022).

Modern supply chains are abundant with SMEs, but they are often unable to reap the same benefits as bigger organizations. They might even weigh down a supply chain and lower its efficacy. All the small business problems that SMEs carry can lower a supply chain's overall ability to handle disruptions and be operationally flexible to accommodate change. The security and data risks of these "weakest links" raise the total risk in the supply chain and the total costs for all members (Lai et al., 2021). The complex demands on SMEs to compete in an ever-changing global marketplace can be taxing and strenuous. Additionally, the digital maturity levels of partners generally limit an SME's ability to properly integrate with its vertical and horizontal partners in the supply chain through any penetrating technologies (Ethem & Susanne, 2019). All supply chain interactions rely on trust enabled through previous history, industry credit, or third-party verification, and trust issues pose barriers to supply chain integration. A lack of transparency with SMEs often lowers their ability to benefit from trust transmission and endorsement from other supply chain players like banks and large enterprises (Jiang et al., 2022), and the whole network operates less efficiently.

General inefficiencies can be found throughout the processes of SMEs. Informal or antiquated inventory, procurement, and accounting practices that have not been kept current, cause problems (Zheng et al., 2022). Communication throughout these processes and with the business environment is also often inefficient, informal, and redundant. Insufficient information needs to be clarification for overlapping stakeholders. Logistical operations for small businesses can be particularly wasteful as products and information are lost in the shuffle between multiple locations and ERP systems. Logistics chains typically have participants with low levels of trust between them (Wu & Zhang, 2022). SMEs suffer from low bargaining leverage with logistics providers since SMEs typically do not provide the same volume as large enterprises.

Two specific but impactful SME inefficiency problems that this study demonstrated and analyzed in depth are insufficient data collection and inefficient inventory policy. Proper data capture, storage, analysis, and use is paramount for modern businesses to function and survive. Companies need to have clear strategies and means to accomplish these actions. Constant streams of highly accurate data are crucial for forecasting, production, strategy, and executive level decision making. Other records like material lot numbers and analysis data have high importance to quality monitoring systems. Due to resource constraints and other problems previously mentioned, SMEs suffer from poor data quality and reach. Data storage is costly and time consuming due to formatting, identification, and availability challenges (Anne et al., 2016). Support tools used by larger companies can be unsuitable because they are too expensive or too complex (Teerasoponpong & Sopadang, 2022). Data often originates from static sources or remains static throughout its use (Hildebrand & Schneider, 2020). Data is typically entered into multiple systems by humans using many methods, resulting in uneven quality and uncorrected errors (Jiang et al., 2022). Some internal, centralized databases may exist with overlap in the types of data they track yet feature asymmetry and lack of consensus persists. Processes that capture data feature blind spots and gaps representing failure to capture the necessary data. The data may not be requested from the source, available from the source, or deemed necessary to capture and save. Captured data is sometimes lost in movement between ERP systems or between firms. Multiple ERP systems with poor interoperability plus a lack of supply chain integration can cause insufficient transparency and traceability. Without these features, single data points can be manipulated or otherwise wrongfully influenced without the proper means to track down the root cause of failure.

A study done by Anne et al. (2016) collected survey data from 152 SMEs of varying sizes in Ireland to assess the readiness of their collected data for use in computerized event simulations, which rely on large amounts of high-quality data. It was found that smaller SMEs captured less data than larger SMEs and that data collection quantities did increase as a company grew in age. There was no correlation between the age of the company and type of data they collected. The predominant collection method for the SMEs was a combination of manual record keeping and some utilization of IT systems. Automated collection technique utilization did increase with the size of the SME. Overall, only a small portion of the surveyed SMEs had data sufficient for use in simulations. The SMEs did not collect as much data as larger enterprises and were less ready to adapt sophisticated data collection methods.
Poor and insufficient data makes it a challenge to manage a proper, efficient strategy when it comes to inventory. In manufacturing, techniques of inventory management like Just-In-Time (JIT) rely on accurate data from suppliers, the stockroom, production, and procurement departments. Without an effective inventory policy, an SME may use inefficient and costly techniques to safeguard against inventory related risks. They are more likely to make decisions based on intuition or experience rather than data (Teerasoponpong & Sopadang, 2022). SMEs typically use many suppliers for the same materials to hedge against stock-outs or other trust and relationship issues (Lai et al., 2021). As with most companies, sourcing and inventory decisions have massive effects on the financial performance of SMEs.

Another guard against stock outs is excessive safety stock. High inventory levels are often used to conceal other problems and deficiencies in operations. The problem with holding more stock than required is that it is an expensive practice. There are facility, maintenance, and tax costs that are directly related to stock levels. There are also risks of products and materials being damaged, destroyed, lost, or expiring in storage. Market demand for products can also evaporate with consumer taste changes or technological breakthroughs. Additional risks include threats to health and safety if the materials are hazardous. SMEs may hold excessive inventory when forced into large minimum buys by suppliers which they do not have proper leverage over, or the opportunity to switch away from. A lack of data and resources to maintain an efficient inventory system is a major hindrance for SMEs while they face market uncertainties. There is research to show that a data-powered inventory model could provide accurate recommendations for an SME and realize inventory cost savings.

LITHIUM-ION BATTERIES

Battery technology has been slowly progressing for decades as technological innovation and chemical formulation advancements have increased energy density and reduced costs. Lithium-Ion Batteries (LIBs) have emerged as the dominant rechargeable batteries sold on the market due to their lifespan, performance, and effectiveness in powering everyday portable electronics and electric vehicles. Global trends towards renewable energy and the shift away from fossil fuel consumption has greatly increased demand and production for these battery types for different applications. Estimates claim that by as early as 2030, renewable energies will be cheaper and more viable than fossil fuel energy (Koyamparambath et al., 2022).

Battery cells are typically made of the same components, i.e., an anode, a cathode, current collectors, a separator, and electrolyte. Electrolyte carries positively charged lithium ions from the anode to the cathode through the separator membrane, freeing electrons in the anode that get stored in a current collector and create a charge. This allows an electrical current to flow from the positive collector, through an electronic device, and out to the negative collector when the positive and negative terminals are engaged. In the charging process, a similar chemical reaction happens but in reverse. LIBs have proven to be very effective in storing electrical energy for future deployment and have properties of high energy density and power density.

LIBs contain several finite earth elements whose physical properties are ideal for this electro-chemical process to work. The anode is typically composed of graphite, while the cathode is made up of lithium, cobalt, aluminum, nickel, and manganese or ferrous oxides. Copper and aluminum are normally used as the current collectors. The electrolyte is composed of lithium salts, various solvents, and additives. Separators are usually a synthetic resin such as polyethylene (PE) and polypropylene (PP) (Koyamparambath et al., 2022). Additional elements found in other parts of LIBs are nickel, silicon, and zinc. Carbon fiber, plastics, or stainless steel are commonly used to encase the batteries into individual cells, modules, or packs depending on the quantity of cells.

The elements used in LIB production come from various mines or recycled sources all over the world and create a very complex supply chain. They are not evenly distributed; minerals such as cobalt, lithium and graphite are concentrated in relatively few countries (Antônio et al., 2022). Australia and South American countries like Chile are major lithium producers. Lithium products are derived from lithium brine or lithium ore. Countries such as Canada and the United States have an abundance of lithium that they have yet to significantly mine. Most cobalt is sourced in the Democratic Republic of Congo. China controls and owns the production of many “rare earth minerals”, magnesium, and graphite. They control 80% of global graphite production. Canada is a leading producer of zinc and nickel. Mexico and South American countries like Peru and Chile have the largest copper mines in the world, but China has the highest processing and trade volumes of imports and exports. No one country dominates the copper supply chain; Canada, U.S, China, Germany, India, and Japan are all massive consumers and exporters of the material (Li et al., 2021). Countries like Australia and Canada are currently expanding mining operations to access massive untapped resources in many of these elements (Koyamparambath et al., 2022).

As demand continues to increase for finite elements used in LIBs, the industry will change as countries maneuver for control of material sources and stronger supply chain links. Aerospace, defense, energy, and electronic industries around the world all recognize the need to secure their supplies of LIBs and their component materials in the coming decades (Mugurusi & Ahishakiye, 2022). Parallels can be made to the global race for countries to secure oil supplies in the last two centuries. Countries such as Japan and the US are investing heavily in the development of material substitutes for LIB production, while other nations are working on strengthening relations with material-rich countries (Koyamparambath et al., 2022). Some materials such as graphite have acceptable synthetic substitutes, but most do not. The U.S. has recognized 35 LIB raw materials as critical to its economy, of which they are 100% import reliant on 14. In 2017, the U.S. president issued an executive order to identify new domestic and foreign sources of critical minerals and to strengthen and streamline all supply chain activity and infrastructures. An effort to “nearshore” more LIB production and supply sources on U.S soil has become an urgent national security measure (Exec. Order No. 13,817, 2017).

The LIB raw material supply chain network continues to grow in complexity as the volume of materials traded continues to increase. For many of the materials, scaling up the supply chain has been a massive challenge. Additionally, material source tracing has proven to be very difficult. Governments and countries are investing heavily in combating supply chain risks. One of the best ways to do so is to diversify the materials used and diversify supply sources employed. Investments in research for breakthrough technology in LIBs and their chemical formulations are being used to find cheaper and more available alternative elements and metals. By cultivating more alternate materials and alternate suppliers, LIB manufacturers can ensure their supply chains are more responsive to disruptions and are ready for production changes. There are also significant efforts and interests in developing more efficient recycling networks for these materials that would otherwise be destined for landfills at their end of use.

The obstacles and challenges for manufacturers to secure LIB raw materials are substantial and continuing to grow. Many powerful countries and large corporations are hoarding away materials or buying entire mines, refineries, and supply chain partners. Smaller manufacturers cannot compete with this buying power and have to take bigger risks to secure their own stocks. Massive changes in supply and demand have also caused prices for the elements and compounds used to fluctuate and remain unstable. Production forecasting and inventory planning become difficult to manage. The international nature of the supply chain networks also brings about substantial legal and political friction that act as trade barriers and bottlenecks (Antônio et al., 2022).

Significant issues exist surrounding the mining practices used for these materials in certain countries. Human rights violations like child labor and inhumane working conditions have been confirmed in countries like the Democratic Republic of Congo. For these reasons and other challenges surrounding the Cobalt they export, there has been a push to develop battery cell formulations that use less of the material. Efforts to ensure the materials are "conflict free" are sometimes more theoretical than what can be realistically guaranteed. Environmental concerns have also been raised at mines around the world. The practices are environmentally taxing, produce carbon, destroy local ecosystems, produce dangerous bi-products, and deplete finite resources. In the U.S, The Environmental Protection Agency has banned many types of environmentally damaging mining practices. This has allowed countries like China, who have less stringent labor and environmental laws, to fill the void and capitalize (Koyamparambath et al., 2022).

While there are efforts to recycle LIBs and the elements inside them, much of the material still ends up in incinerators or landfills. The challenge is to make it cost effective for a company to source recycled material over virgin material from the ground. Efforts to create a circular economy for LIB materials have not yet taken off. Materials like copper do have established recycling networks, but many manufacturers require a specific grade in production. Lithium recycling rates are very low because of the expense and the material degradation. Even when a LIB cell is recycled, other elements are more valuable and lithium material is discarded (Shao & Jin, 2020).

For many LIB manufacturers, the grade and quality of the elements used is very important. Battery performance optimization is highly dependent on high quality and specifically graded material. Material is usually tested to verify certain properties before being used in production. As demand for raw materials remains very high, counterfeits naturally make their way into supply chains because it is profitable. Material is fraudulently passed off as higher grade or as coming from non-conflict sources, among other things. Poor quality material has expensive consequences for manufacturers and is difficult to track and trace (Antônio et al., 2022).

Given the sources and material properties, transit and storage problems exist in supply chains for these materials. Many elements are hazardous materials that have more expensive logistical processes. Special facilities are often required for safe storage and to preserve shelf life. Materials like copper oxidize over time. Lithium in its pure form is unstable around moisture and has shelf-life considerations in its powder and salt forms. Most battery cells and materials need to be stored in extremely dry conditions or they can be ruined. As the global supply chain for these materials continues to evolve, more regulations will follow to govern the extraction, trade, purchase, and storage of all materials. Additional requirements for manufacturers to verify conflict-free sources for material, track carbon footprints, and track lifecycle of LIBs may very well come (Antônio et al., 2022).

DATA

Data was gathered from a U.S. based SME manufacturer producing high performance Lithium-Ion Batteries customized for specific uses and applications. These batteries are produced in relatively small lot sizes and rigorously tested to meet tight-tolerance design specifications and performance capabilities. They use a "pull" production method to pull inventory through the system to fulfill orders as they are placed. Access was authorized to their ERP system and plugin software tools, along with their internal server network data. The conditions for access required that the company name, specific product names, and supplier/vendor names be omitted from publication. To honor this request, certain names were changed, omitted, or abbreviated to conceal the identity. This had no effect on the analysis or conclusions. The data gathered for analysis was Manufacturing Order (MO) data, Purchase Order (PO) data, and on-hand inventory data. Manufacturing Order data shows the records of raw material or work in process inventory being consumed in the production of another work in process part or a finished product. For example, an MO record may show a "lid" part and a "can" part being used to create a "case" part, plus all the associated data points about the materials and the transaction. Due to the number of records, only 2 years of MO data could be pulled from the ERP. The time frame was May 2020 to May 2022. A longer time frame would not help this study anyway because battery cell designs and the materials used change too often to see significant long-term trends. Purchase Order data was accessed to show where materials were purchased from, their unit prices, lead times, and other useful data. Inventory records were used to determine how much of each part was stocked on-hand at the time of this study, the units of measure, the dollar value of the parts/materials, and other data. All data used in this study is stored in a Microsoft Excel spreadsheet and available.

The data was cleaned and standardized to prepare for analysis. Much of the data was siloed in different parts of the ERP system, reports generated by the plugin tools, or stored on the network server. It took significant effort to compile all the data in one location. From there, units of measures were standardized, specifically GM and KG. Next, considerations had to be made for parts that had their Part ID changed over the years for various reasons. The exact same physical part may have been stocked or consumed on an MO under multiple Part ID numbers, so manual summations had to be used. The data was then reviewed by an employee with expertise on company operations to clean any noticeable errors from the manual entry nature of the values. With the data scrubbed, Pivot Tables were used to make summations of all the different revision levels used on a given part for MOs and inventory records. For various reasons, the same part appears as different revision levels (A, B, C, D, etc.) over time, which splinters the data. The same part can also currently be stocked under different revision levels and in different inventory locations, so the ERP treats them differently in generated reports. For this study, the grand totals without revision level designation were important to sum up so they could be used in calculations.

A summary spreadsheet was created on Microsoft Excel by pulling all the data together using the "VLookup" function. All materials used in production in the last two years appear here with their usage, procurement history, and inventory statistics summarized. An important piece of information is noted next to the Part ID number, which is a Make/Buy designation. This is an important filter used in the Analysis section to differentiate between parts that are made in house (work in process inventory), and raw materials that are purchased from an outside source. Another designation of “Critical Material” was manually added to qualifying parts as a yellow highlight on the Part ID. This designation was created by the SME for parts that the company deemed to require extra supply chain attention because they are essential to support continued production of cell quantities in various chemistries. These items may have supply, lead time, and procurement challenges associated with foreign vendors or a sole source supplier that would require a significant qualification effort for product substitution. For items with both the “Buy” designation and the “Critical” material designation, manual data regarding the primary element/compound in the material was added, along with a calculation for lead time on last purchase (days difference between PO order date and PO receipt date). These manual data additions were only made for the Buy and Critical designated items due to the time constraints of the study and scope of the analysis.

With all the data summarized, calculations could be made for each material for the sum consumed per year by the SME, the number of years supply on hand, the inventory turnover ratio, the dollar value used per year, the dollar value on hand, and the lead time in days. Once again, lead time statistics were only collected on the critical and buy type materials due to time constraints and scope of the study. Descriptive statistics summaries were drawn for all the Buy materials and all the Critical materials so conclusions on inventory practice efficiency could be formed. ABC Analysis’ based on the dollar value of the material consumed per year was conducted for all the Buy materials and all the Critical materials. Several materials were qualified for further analysis on a Scatter Plot based on their annual consumption dollar value and years of supply on hand. These analysis efforts isolated two materials that motivated further quantitative analysis into the level of upstream data known and available to the company. For these two materials, supplier history and material tracing data were pulled from the ERP system, network database, vendor websites, and through vendor requests. Data was compiled in a table and a Fishbone Analysis was conducted to understand why any data gaps may exist. All data material sources used in this compilation is available upon request.

MODELS OF ANALYSIS

The models in this section were used to explore the current inventory situations for all Buy materials, and all Critical materials. These are not mutually exclusive, and all the Critical materials are also Buy materials. Various criteria were used to limit all observed materials down to two that are expected to be of high importance in securing in the coming years. All known data on the sources of these materials was gathered to discover data gaps in traceability, and further analysis was done to explore why those gaps may exist at this time.

125 Buy parts, or parts and materials purchased from external sources, were used on manufacturing orders from May 2020 to May 2022. A per year average was calculated so data analysis could be done on an annual basis. Using the last unit cost available and the current on hand inventory data, the dollar value of manufacturing order material used per year and the dollar value of the inventory on hand could be calculated for every material. By dividing the current inventory quantity (as of May 2022) on hand by the annual amount of material consumed, the number of years’ worth of supply on hand was calculated. The inverse of this number, or 1/Years’ Supply, was used as an annual inventory turnover ratio in this model. This gives an idea of usage rates per year relative to the stock kept on hand. A starting inventory level and ending inventory level for the period of 1 year was not available, so an average inventory could not be established. The calculated averages and other descriptive statistics for

all 125 Buy parts are displayed in Figure 1. Of the 125 Buy materials, 22 are also designated as Critical by the company per their standards referenced in the Data section of this study. Along with calculations of number of years supply and inventory turn ratio, lead time in days was also calculated for each critical material. To do so, order and receiving data was manually pulled from the ERP system, but only for these materials due to time constraints of this study. Doing so for all of the Buy materials could be done in future studies. The calculated averages and other descriptive statistics for all 22 Critical parts are also displayed in Figure 1.

"Buy" Materials	Supply on Hand (Years worth)	Inventory Turnover Ratio
Mean	53.42102802	0.451005665
Standard Error	12.94654629	0.098522391
Median	9.625	0.103896104
Mode	#N/A	#N/A
Standard Deviation	144.7467879	1.101513816
Sample Variance	20951.63262	1.213332687
Kurtosis	32.92656222	23.86134442
Skewness	5.383874771	4.57452111
Range	1146.949406	7.953673672
Minimum	0.125714286	0.000871782
Maximum	1147.07512	7.954545455
Sum	6677.628503	56.37570811
Count	125	125

"Critical" Materials	Supply on Hand (Years worth)	Inventory Turnover Ratio	Last Lead Time (Days)
Mean	36.72040891	0.441897241	56.59090909
Standard Error	15.98842886	0.182567656	10.08355978
Median	7.155432535	0.140876149	44.5
Mode	#N/A	#N/A	30
Standard Deviation	74.99237872	0.856318212	47.29608772
Sample Variance	5623.856866	0.733280879	2236.919913
Kurtosis	12.79983371	7.999774758	3.178637555
Skewness	3.395542308	2.81932534	1.638937798
Range	335.700944	3.509586496	198
Minimum	0.284692418	0.002976318	4
Maximum	335.9856364	3.512562814	202
Sum	807.8489961	9.721739292	1245
Count	22	22	22

Figure 1: Descriptive Statistics Table

An ABC Analysis of the Buy Materials was conducted in Figure 2 based on dollar value used per year per material. The ABC Analysis sorted materials by total dollar value per year and calculated the percentages each contributed to the total dollar value of all Buy materials used. The cumulative running percentage was calculated in the next row and was the basis of the ABC designations. Only "A" and "B" materials are shown in the figure due to size limitations. "A" materials are the most important to the company. They make up the top 80 percent of all dollar value of consumed material, yet typically make up about 20 percent of the population. In Figure 3, the Critical materials were isolated while all other Buy materials were filtered out, and an ABC analysis was again under the same conditions.

"Buy" Materials				
Description	**Dollar Value used per Year**	**% Total DV**	**Cumulative %**	**ABC Designation**
LID ASSEMBLY	$ 39,257.13	23.42%	23.42%	A
CASE	$ 10,556.40	6.30%	29.72%	A
N-METHYL PYROLIDINONE	$ 9,803.86	5.85%	35.57%	A
TITANIUM FOIL 12 MICRON x 200 MM	$ 9,491.11	5.66%	41.23%	A
LID ASSEMBLY	$ 8,855.10	5.28%	46.52%	A
MODEL 3A LID ASSEMBLY	$ 7,718.70	4.61%	51.12%	A
COPPER FOIL, 8 MICRON x 250 MICRON	$ 6,156.98	3.67%	54.80%	A
MODEL 6 CASE	$ 5,616.63	3.35%	58.15%	A
HIGH VOLTAGE LiCoO2	$ 5,307.90	3.17%	61.31%	A
COPPER FOIL 10 MICRONS x 250MM	$ 4,960.09	2.96%	64.27%	A
LID ASSEMBLY	$ 3,529.84	2.11%	66.38%	A
ALUMINUM FOIL, 15 MICRONS x 250 MM WIDE, Al 1085	$ 3,281.70	1.96%	68.34%	A
POSITIVE EXIT TRANSITION TAB	$ 3,281.25	1.96%	70.30%	A
CONTINUOUS POSITIVE COATING DS	$ 3,034.00	1.81%	72.11%	A
CPreme G8 GRAPHITE POWDER	$ 2,996.28	1.79%	73.89%	A
CASE	$ 2,572.95	1.54%	75.43%	A
LITHIUM HEXAFLUOROPHOSPHATE	$ 2,539.59	1.52%	76.94%	A
ALUMINUM LAMINATED FILM	$ 2,377.45	1.42%	78.36%	A
GRAPHITE POWDER	$ 2,150.26	1.28%	79.65%	A
METHYLENE METHANEDISULFONATE (MMDS)	$ 2,137.50	1.28%	80.92%	B
POSITIVE EXIT TAB (ALUMINUM)	$ 1,822.50	1.09%	82.01%	B
Cu/Ni TAB,0.1 X 5 X 70-PP0.1 X 5 X 9 w/DNP BLK FILM	$ 1,820.00	1.09%	83.09%	B
SC37-0500 Silicone Cap	$ 1,729.60	1.03%	84.13%	B
GASKET	$ 1,392.18	0.83%	84.96%	B
H1612 SEPARATOR, 150.0 MM	$ 1,340.64	0.80%	85.76%	B
ETHYLENE CARBONATE	$ 1,338.95	0.80%	86.56%	B
NEGATIVE EXIT TAB (NICKEL-PLATED COPPER)	$ 1,316.75	0.79%	87.34%	B
MCMB 2528 CARBON	$ 1,234.62	0.74%	88.08%	B
CONTINUOUS NEGATIVE COATING DS	$ 1,221.37	0.73%	88.81%	B
MODEL 5 CASE	$ 1,031.55	0.62%	89.42%	B
ALUMINUM LAMINATED FILM (D: EL40H(III)), 113 um	$ 1,028.93	0.61%	90.04%	B
LiFePO4 POSITIVE ELECTRODE COATING	$ 829.725	0.50%	90.53%	B
POSITIVE EXIT TRANSITION TAB	$ 828.75	0.49%	91.03%	B
G5 NEGATIVE ELECTRODE COATING, DOUBLED SIDED	$ 806.46	0.48%	91.51%	B
1,3 PROPANE SULTONE	$ 787.34	0.47%	91.98%	B
SLURRY FEED SYSTEM FILTER, ROKI TECHNO 125L-HC-200AD	$ 757.50	0.45%	92.43%	B
ALUMINUM FOIL 20 MICRONS,THK x 300MM W	$ 700.84	0.42%	92.85%	B
ELECTRO-DEPOSITED COPPER FOIL, 10 MICRONS	$ 640.54	0.38%	93.23%	B
CASE	$ 606.04	0.36%	93.59%	B
CPreme GRAPHITE POWDER	$ 574.89	0.34%	93.93%	B
DIETHYL CARBONATE	$ 518.10	0.31%	94.24%	B
ETHYL METHYL CARBONATE	$ 516.72	0.31%	94.55%	B
TODA NAT-9152 NCA	$ 462.25	0.28%	94.83%	B

Figure 2: ABC Analysis of “Buy” Material

To identify and isolate the Critical and Buy designated material that will be of special importance to the manufacturer in the coming years, a narrower targeting approach was taken. As demand for LIB material increases, this SME will need to prioritize which materials they most urgently need to secure. The Critical materials were narrowed down to focus on just materials that have under 10 years of stock on hand, which includes 14 of the 22 critical materials. The idea here is that these materials are used often, and additional supply will be required soon. By focusing on items that will have to be reordered in the next 10 years (near term), major data outliers were eliminated that would have affected the Scatter Plot in Figure 4 below. This Scatter Plot was created to compare material dollar value consumed per year against years’ supply on hand.

"Critical" Materials				
Description	**Dollar Value used per Year**	**% Total DV**	**Cumulative %**	**ABC Designation**
N-METHYL PYROLIDINONE	$ 9,803.86	17.83%	17.83%	A
TITANIUM FOIL 12 MICRON x 200 MM	$ 9,491.11	17.26%	35.10%	A
COPPER FOIL, 8 MICRON x 250 MICRON	$ 6,156.98	11.20%	46.30%	A
HIGH VOLTAGE LiCoO2	$ 5,307.90	9.66%	55.95%	A
COPPER FOIL 10 MICRONS x 250MM	$ 4,960.09	9.02%	64.98%	A
ALUMINUM FOIL, 15 MICRONS x 250 MM WIDE, Al 1085	$ 3,281.70	5.97%	70.95%	A
POSITIVE EXIT TRANSITION TAB	$ 3,281.25	5.97%	76.92%	A
LITHIUM HEXAFLUOROPHOSPHATE	$ 2,539.59	4.62%	81.54%	B
ALUMINUM LAMINATED FILM (D-EL35H(III)	$ 2,377.45	4.32%	85.86%	B
GRAPHITE POWDER	$ 2,150.26	3.91%	89.77%	B
NEGATIVE EXIT TAB (NICKEL-PLATED COPPER)	$ 1,316.75	2.40%	92.17%	B
MCMB 2528 CARBON	$ 1,234.62	2.25%	94.41%	B
ALUMINUM FOIL 20 MICRONS,THK x 300MM W	$ 700.84	1.27%	95.69%	C
TODA NAT-9152 NCA	$ 462.25	0.84%	96.53%	C
POLYVINYLIDENE FLUORIDE (PVDF)	$ 452.35	0.82%	97.35%	C
VAPOR GROWN GRAPHITE TUBE (VGGT)	$ 411.16	0.75%	98.10%	C
HSV1810 PVdF	$ 329.63	0.60%	98.70%	C
PRE TAPED NI PLATED COPPER BATTERY TAB, 5 MM x 57 MM	$ 234.50	0.43%	99.13%	C
PROPYLENE CARBONATE	$ 195.70	0.36%	99.48%	C
PRE TAPED ALUMINUM BATTERY TAB, 5 MM x 57 MM	$ 180.00	0.33%	99.81%	C
SFG-6L GRAPHITE	$ 84.78	0.15%	99.96%	C
KS-6L GRAPHITE	$ 20.44	0.04%	100.00%	C

Figure 3: ABC Analysis of "Critical" Material

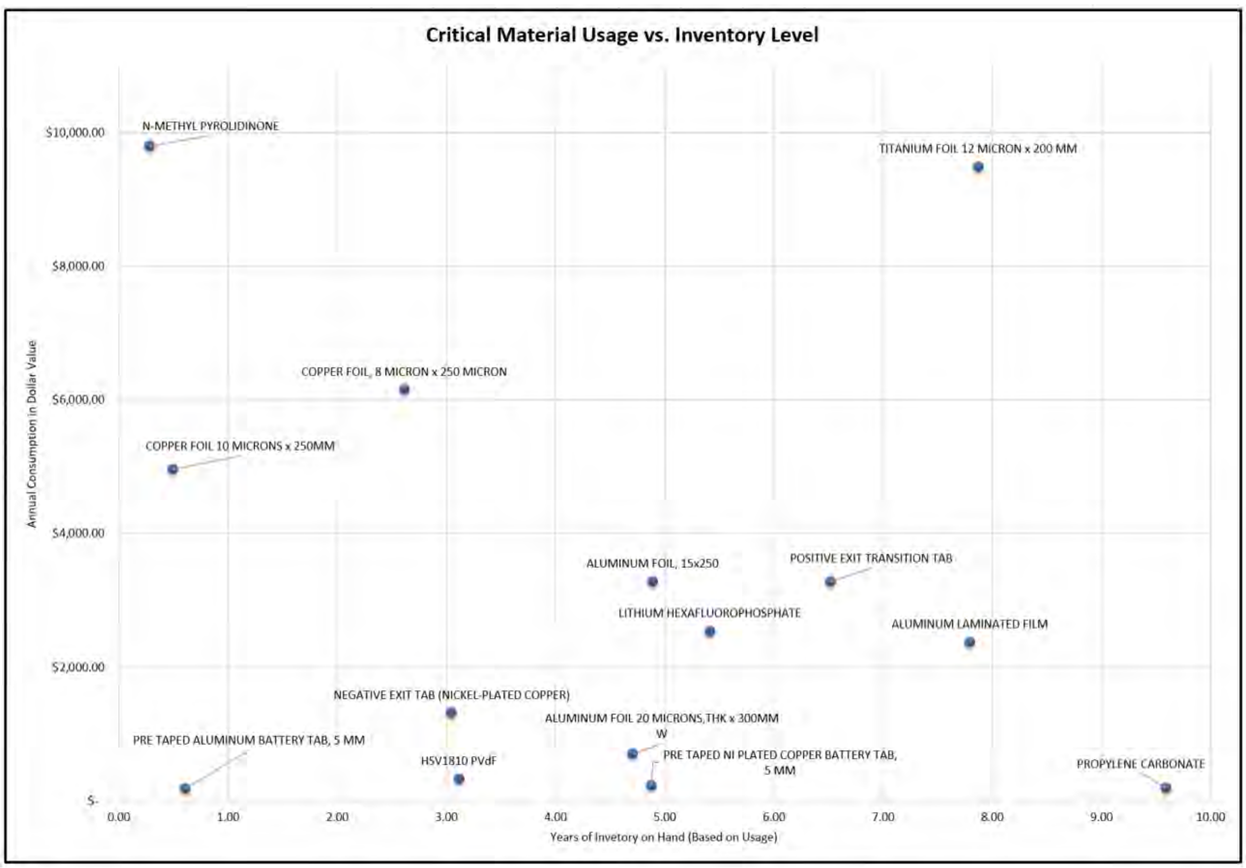


Figure 4: Critical Material Scatterplot

This scatterplot visualization physically sorted the qualified critical materials into different areas of importance. Materials that cost the most per year to satisfy production are higher on the Y axis, while materials that are projected to run out sooner are lower (left) on the X axis. For these reasons, the upper left quadrant displays the materials requiring the most urgent attention. They are of high importance to the company based on annual expenses and the relatively low current stock levels. The upper left most material, N-Methyl Pyrrolidinone (NMP) is the most prominent example of this. The other materials in this quadrant are Copper Foils. Additional materials such as Aluminum Foil and Lithium Hexafluorophosphate (LiPF6) are further away from the top left but still important due to their central position. Annual usage is above $2,000 in these materials and there is less than 6 years of stock, which can only be said for 5 materials represented here. Materials that were very low on the Y axis and very high (right) on the X axis were subsequently disqualified from future evaluation in this paper.

Using Figures 1-4, two materials were selected for analysis on data traceability, data sufficiency, and presence of gaps in sourcing data. The criteria used was that the material should be an A material on both ABC analysis, appear near the target quadrant on the scatterplot, have shelf-life considerations, and primarily contain a finite earth element with a complex global supply chain. The material NMP was ruled out for not primarily containing an element of target; it is a very commercially available chemical solvent. Two Copper Foil parts qualified, so one was selected as the category representative since the only difference was the thickness. The 10-micron thick version was selected since on hand inventory was less than the other. The other material chosen for the analysis was Lithium Hexafluorophosphate (LiPF6), also known as Lithium Salt. It is an A tier Buy material but just missed the cut as an A tier Critical material by 1.54 percentile points in Figure 3. It was still selected because it was near center on the Scatter Plot in Figure 4 and is prone to expiring. Like the Copper Foil part, this material is made from a finite element and purchased to a very specific grade, which presents challenges in finding substitute parts and suppliers. Copper and Lithium parts are also very representative of typical LIB materials universally used by all manufacturers. Competition for these materials will undoubtedly present challenges for similar SMEs in the coming years.

All known sourcing data was compiled for the Copper Foil and Lithium Salt parts into the Data Gap Visualization Table in Figure 5. This table traces all known supply chain data for each material, from internal company identification to direct suppliers, to all known parties back to the originating mine(s). The table aimed to identify the material’s various part numbers, logistical and procurement considerations, supplier geographical locations, and to map the supply chain parties that add value. Internal company experts were consulted to discover all relevant information known to the company. Missing or vague pieces of tracing data are highlighted in gray. As the table is read from top to bottom, the information sources get further away in proximity to the company. This increases the difficulty in data access, visibility, and capture.

Data Proximity	Information Source	Name	Copper Foil		Lithium Salt (LiPF6)
Source	At Use	Part Description	COPPER FOIL 10 MICRONS x 250MM		LITHIUM HEXAFLUOROPHOSPHATE
		Facility Part Number	MMMP1738		CBA102
		Primary Element	Copper		Lithium
		Process Use	Anode Current Collector		Electroyte Ingredient
		Former Suppliers Names	"Supplier PM"		"Supplier PM"
		Reason of Discontinuation	Company Aquired by current supplier		Company Aquired by current supplier
Near	Direct Supplier	Direct Supplier Name	"Supplier DA"	"Supplier U"	"Supplier U"
		Direct Supplier Type	Foil Producer/Manufacturer	Importer/ Reseller	Importer/ Reseller
		Direct Supplier Part Number	TLB-PLSP	NC-WS-10	
		Direct Supplier Location	Camden, SC, USA	Lyndhurst, NJ, USA	Lyndhurst, NJ, USA
		Direct Supplier Lead Time	65 days	44 days	36 days
		Direct Supplier Logistics Info	Ground Freight	Ground Freight	Ground Freight
		Terms of Purchase	100% in advance of shipment	Net 60	Net 60
		Methods of Purchase	PDF Quote and PO via email	PDF Quote and PO via email	PDF Quote and PO via email
		Additional Notes	Operation purchased from "Supplier OM", a subsidiery of "Supplier MK"		
Near	Intermediary Supplier	Indirect Supplier Name	"Supplier ND"	"Supplier F"	"Supplier K"
		Indirect Supplier Type	Material Exporter	Foil Producer/Manufacturer	
		Indirect Supplier Part Number	TLB-PLSP		
		Indirect Supplier Location	Tokyo, Japan	Nikko City, Tochigi Prefecture, Japan	Osaka, Japan
		Indirect Supplier Lead Time			
		Indirect Supplier Logistics Info	Air Freight	Air Freight	Air Freight
		Additional Notes			
Far	Other Intermediaries	Indirect Supplier Name	Unknown if any exist	Unknown if any exist	Unknown if any exist
		Indirect Supplier Type			
		Indirect Supplier Part Number			
		Indirect Supplier Location			
		Indirect Supplier Lead Time			
		Indirect Supplier Logistics Info			
		Additional Notes		"Supplier F"'s comments: "Copper from recycled material. Unknown origins of the ore"	
Far	Refiner	Refiner/Company Name	"Supplier MK"		
		Refiner Country(s)	Japan		
		Additional Notes	Suggested as the likely the Refiner/Smelter, but can't be confirmed		
Distant	Mine	Mine/Company Name			
		Mine Country(s)			
		Additional Notes			

Figure 5: Data Gap Visualization Table

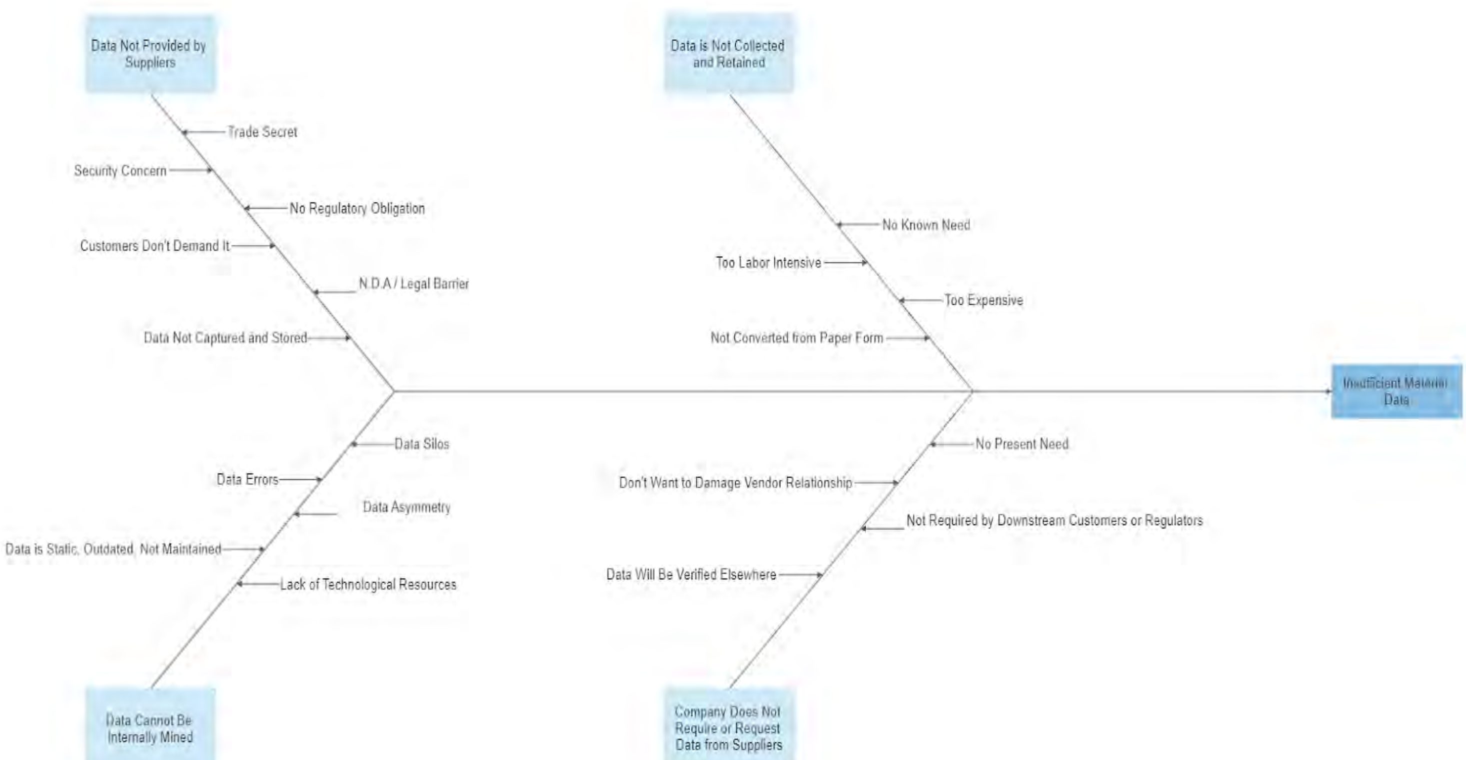


Figure 6: Fishbone Diagram

Tracing information about the sources of each product may be unavailable in the company ERP and/or internal network for various reasons. Each of them is explored in the Fishbone Diagram in Figure 6. The Fishbone Diagram was used to do a qualitative root cause analysis as to the core reasons why this information is unknown. It was determined that the missing data was either not requested by the company, not provided by the upstream suppliers, not collected if it was supplied, or not able to be mined out of the internal systems. Each of these potential root causes have contributing factors as to why they may occur, which are also displayed.

RESULTS AND RECOMMENDATIONS

The models in the previous section were created in succession. A master spreadsheet was compiled, and calculations were computed that were used to create the Descriptive Statistics Table in Figure 1 and the ABC Analysis' Figures 2 and 3. Off of these models, the Critical Material Scatter Plot in Figure 4 was built, which then spawned the Data Gap Visualization Table in Figure 5 and the Fishbone Diagram in Figure 6. In building these models, the original master spreadsheet and the Data Gap Visualization Table required the most manual data collection labor. Data available in the company ERP system and other network locations had many quality and integrity issues. The data also had to be manually pulled out of multiple siloed locations. Errors had to be reviewed for correction by an internal expert. Performance indicator statistics in Figure 1 had to be calculated in the master spreadsheet; they were not previously readily available or maintained for use by any company analysts.

The Data Gap Visualization Table in Figure 5 provides indication of the blind spots the company has when it comes to tracing these materials back to a source location. This provides an indication about how much they readily know about their supply chain for 2 products that are heavily used, critical to production, and derive from finite elements. This sample can give an indication of the level of data collected on other Critical materials. At the material's use location in the manufacturing facility, much data is known about each part's identifying characteristics. Internal testing can verify the material quality and the quantity on hand is known and verified by regular inventory counts. Both Copper Foil and Lithium Salt were previously sourced at a vendor that was bought out by a current supplier. Copper Foil has two approved, qualified sources in Supplier DA and Supplier U. Much is known about these companies and their material processes from vendor onboarding and material qualification efforts done at some point in time by the SME manufacturer. Procurement and logistics procedures have a known history; Supplier U extends a line of credit and has a shorter lead time than Supplier DA. The same level of data exists for the direct supply of Lithium Salt, which is also sourced at Supplier U, except for their internal part number for the product.

Moving up the supply chain to intermediary suppliers, much less data is available to the company. The 3 suppliers used for these 2 materials all have exporters or producers in Japan, but limited information is known about the material part numbers/names or lead times to the United States. There are only potential suggestions as to the origins of the elements used in the Copper Foil and the Lithium Salt. These suggestions were provided by the direct suppliers themselves when requested, this was not data stored in company databases. No data regarding source mines could be found for either product's supply chain networks.

The data gaps and lack of transparency that exist in an upstream analysis of the company's materials can result in problems and unforeseen risks for the company. Root Cause Analysis tracing cannot be conducted up the supply chain on quality issues observed by the battery manufacturer. The material source can't be truly verified as "conflict free" or verified for other contractual or regulatory purposes. This data can't be passed on to a customer, so they lose out on a competitive advantage. These data gaps also make it much harder to understand the impact of global events or logistical delays. Based on what is known about Copper and Lithium mining, it is unlikely that these materials originate in Japan except for the potential that Supplier F sources their recycled copper there.

The Fishbone Analysis was conducted to explore all possibilities of why upstream data gaps exist and other reasons why transparency is difficult to obtain. Upstream product data can only be unavailable for four main reasons; suppliers don't provide it, the company does not request it, the company does not collect it, or the company can't mine it out of their systems. The Fishbone Analysis displays the potential contributing factors that go into each of those possibilities. Based on the research done for this study and the opinion of company experts, the 2 materials that were analyzed are likely limited in data transparency by: legal barriers, fear of damaging supplier relationships, and a lack of retention/conversion of the information from paper form. Supplier U noted that an upstream source would not provide additional material sourcing information without a nondisclosure agreement in place. The SME manufacturer also has lower reporting requirements from these vendors since many materials are sole-sourced and they do not want to disqualify any vendors. To preserve the relationship, the manufacturer does not want to overburden the suppliers with extra requirements. Many of these material shipments also arrive from overseas with paper export declarations and other material certificates that are discarded without a known present need or filed away in their paper form. It should be noted though that all the Fishbone contributing factors most likely exist in varying degrees for all materials purchased by the company. Future studies should be done on additional Critical materials to understand other unique barriers to collecting sourcing data and to understand the nature of other suppliers.

The Descriptive Statistics and ABC Analysis figures took a more quantitative look at material usage to understand the efficiency of the inventory policy in place. As noted, much effort was expended to gather the necessary quantitative data, which would be expensive and time consuming to replicate often for all materials used and purchased by the company. Figure 1 demonstrated that both the Buy materials and Critical materials had a very high average amount of supply on hand when based on years of supply, at about 53 years and 36 years respectively. They were influenced by very large outliers and spread out over a very large range, but around half of the Buy materials had over a decade's worth of stock on hand at current usage rates when the data is viewed by part. An extreme example of this was found in one Buy material: Packing Putty. $31.40 worth of the material has been used per year on average, yet there is $36,018.16 worth of it stored in inventory. Minimum purchase quantities that far exceed what is needed may have influenced some of the excessive inventory quantities like this.

Average inventory turnover ratios for both material categories were only .01 apart, but the range was much wider for the Buy materials. At ratios of .45 and .44 for Buy and Critical Materials respectively, inventory turnover is very poor across all materials. A company's goal should be to turnover inventory as often as possible, especially when using materials that have a shelf life. A ratio of 4-7 would be much healthier than turning over the Critical materials at an average of less than once every 2 years. Only 3 of the 22 Critical materials had the equivalent of 1 full inventory turn per year observed. Average lead time was only calculated for the Critical materials but still delivered meaningful statistics. At 56 days on average, these materials can take considerable time to arrive, and the range varies by 198 days. The 48-day standard deviation indicates uncertainty may be high and makes forecasting difficult. Despite this, with lead times that average around 2 months, it is rather excessive to keep multiple decades of stock (on average) on hand and demonstrates inefficient inventory practices.

With decades of stock on hand for both highly used and seldomly used materials, it is easy to say that the company has not embraced any Just-In-Time inventory methods to keep average inventory levels low, or a lean inventory method to eliminate as much waste as possible. This manufacturer is known to use a pull production method to pull inventory from suppliers and through their processes, but it has been demonstrated that they do so with a tremendous buffer of safety stock. Despite its expenses and risks, this buffer is to accommodate unpredictable customer demand, changing supplier conditions, and the risk of loss of material access. The annual price for a material cushion like this, expressed as the annual holding cost, is often estimated at 20% of the material value as a standard. With a $1,524,8889.35 value of Buy materials at the time of this study, the annual holding cost would be $304,977.87.

As previously explained, the Copper Foil and Lithium Salt parts were selected for further qualitative analysis based on their relatively lower stock levels, A tier (or very close) designation on the ABC Analysis tables based on usage, the predominant presence of finite earth elements, shelf-life considerations, and previous identification by the company as Critical. In the Buy Material ABC Analysis in Figure 2, 15.2% of the materials made up the top 80% of all material consumed in dollar value. 24 B materials made up the next 15% of spend cumulatively and represented 19.2% of the population. The remaining 82 materials, the C group, made up the final 5% of dollar value of all materials used, but made up the bulk of the population at 65.6% of the materials. In the Critical Material ABC Analysis in Figure 3, 7 materials made up the most important A designation, 5 made up the B designation, and the remaining 10 materials made up the C designation. It is worth noting that all these A materials were also classified as A materials amongst all 125 Buy parts. Based on the ABC Analysis of Buy materials, the company has done well in categorizing the Critical materials; the majority are A tier in terms of usage. It will be important for the company to continuously recalculate A materials year after year since these are liable to change. Additional data collection on these materials will help with forecasting and company strategy. These are also the more expensive materials and should be reviewed in cost saving efforts.

Sufficient data collection and efficient inventory practices work well in tandem and are greatly influenced by each other. As is the case with many organizational processes, problems with each of them can compound in a vicious cycle. These data and inventory problems are common in many SMEs and have been observed in this case example. Limited data management resources, general digital immaturity, and a lack of integration with suppliers affect data quality and sufficiency. Offline, redundant, paper-based data sharing, and informal communication creates poor data retention into digital forms that can be used in analysis. The limitations of data processing power and of the quality of data that can be captured creates process bottlenecks. As for the inventory practice, symptoms of low bargaining and purchasing power have been observed in the model data. The manufacturer has unique challenges in sourcing highly demanded, finite, and often foreign-based production material. For these reasons, it appears that they have built a buffer inventory that exceeds their needs but safeguards against future supply chain challenges. The nature of these material's markets and the company's aversion to supply risk creates a constraint that prevents lean inventory practices from being used.

Blockchain technology-based interventions can be used to alleviate these bottlenecks and constraints. By participating in a Blockchain-enabled supplier marketplace, this SME would be able to find more qualified suppliers of their commonly purchased and Critical materials and lower the barriers to changing suppliers. In an environment where supplier credibility and performance can be verified by other Blockchain participants, trust can be enabled to do business with unknown parties. The SME manufacturer can also pool resources with other buyers in the marketplace to have access to constant supplier audit data, rather than running one audit at the time during onboarding and relying on this static data going forward. Suppliers that are subject to constant audits and have verifiable contractual performance data that is shared with a network will also be incentivized to perform better and uphold their reputation. This should lower the threat of counterfeit and sub-par quality LIB materials from ending up in their products. With this level of visibility, the SME manufacturer can improve their research and development processes by finding new vendors with new material sources. By accessing more verified data, they can build a vision of what their supply chain looks like geographically to understand how global events might affect them. They can qualify more diverse suppliers for the same production materials and open up more options for interoperability that Blockchain can enable. The effort can also aid the company in finding more domestic sources for certain materials, lowering the risks of using complex global supply chains when possible.

Through Blockchain, materials can move through the supply chain with many levels of data directly tied to them. The SME manufacturer can receive materials with verified testing results, sourcing certificates, and a full chain of logistical custody proof. This can provide the company with a transparent record of all the miners, refiners, and suppliers that handled the material. The producers and middlemen can still securely protect the data they do not want shared by using certain security measures within Blockchain, but the network can still confirm certain pieces of information like country of origin and claims of being "conflict free". Overall, more upstream supplier claims can be trusted with fewer inefficient verification efforts used. Automated processes would also increase the quantity and quality of data collected and transferred. This makes data more suitable for machine learning, storage, and integration into other Industry 4.0 tools and processes, such as IoT devices and smart manufacturing.

With more high-quality data being fed into inventory control procedures, the company can use more efficient policies to help them compete with bigger industry players. By incorporating Blockchain tools like supplier integration and smart contracts, processes redundancy and the inefficiencies of manual data verification can be removed. By participating in a Blockchain verified data ecosystem with other supply chain parties, the SME manufacturer can take advantage of credit penetration and credit radiation effects to obtain better credit terms for purchases. With access to verified supplier data on critical materials, in addition to more supplier options, the company can take steps towards utilizing more efficient inventory processes like JIT and Lean Methodology instead of relying on a large inventory buffer to safeguard them against risk. This in turn will lower their inventory holding costs and improve their bottom line.

It is recommended that in the near term, this company continues efforts to properly identify the most critical materials for production and the suppliers associated as the company grows and global competition for LIB material increases. The company should continue to keep up with modern technology that will help them run efficiently and complete as a small to medium enterprise. They should prepare for Blockchain integration into their industry and supply chains by learning more about the technology and hiring people who understand it throughout their organization. It is important for them to track which partners, ERP service providers, suppliers, and competitors are transitioning towards using the technology and preparing for integration. Blockchain only works, and is strengthened by, many parties collaborating through it. It is recommended that they reach out to their most important vendors to understand their plans to use Blockchain and to request this data from new suppliers during onboarding audits. Larger companies in the industry may already be using Blockchain and if so, the company should investigate how they could possibly get involved in the network.

In the longer term, this company should be ready to integrate Blockchain into their manufacturing processes and ERP systems. They should be aware of any Blockchain middleware software the ERP service is integrating into their platform. For this company, and any SME company, a full cost-benefit analysis of Blockchain integration should be conducted to fully understand the risks and benefits of implementation. Decision makers will have to determine a strategy going forward and the costs of implementation. Investing in the technology, infrastructure, and staff will be very expensive and may not be right for every company.

CONCLUSION

Small and Medium Enterprises face many obstacles to compete with larger companies in this era of globalization. For SME manufacturers, sufficient data collection and inefficient inventory procedures can expose the company to extra risk and expenses. Under the challenging circumstances of securing materials for Lithium-Ion Battery production, these problems were confirmed and analyzed in one case company example. The benefits brought about by Industry 4.0 technologies such as Blockchain provide opportunities to alleviate these problems and help SMEs prepare for growth during increased competition. Trust, transparency, and verification that Blockchain can apply to data and relationships can allow an SME to diversify its supply chain partners and collect higher quality data. This can lower the stress caused by inefficient practices that have been traditionally used to mitigate the associated risks. By joining a network of Blockchain users, companies open themselves to better integration with their business partners, saving the time and money associated with third party verification systems and non-automated processes. This study only looked at one SME and analyzed their processes surrounding relatively few materials, but the opportunity for improvement was still illustrated as an example. Future work can consist of a more in-depth look at all Buy materials to understand more about their inventory transaction histories and gaps in known material source data. Constant reevaluation of “Critical” and “A” designated material can help the company adjust which materials and vendors to focus its efforts on. It is currently early in Blockchain technology’s wide scale adoption but being unprepared will likely be costly to late adopters. It only makes sense to invest in the technology once other supply chain partners are also willing to do so. All SMEs should seek guidance from technological consultants and industry trade unions to understand how they can join any existing information sharing networks enabled by Blockchain.

REFERENCES

Anne Marie, I., James, B., & Byrne, P. J. (2016). Analysis of SME data readiness: a simulation perspective [Journal]. *Journal of Small Business and Enterprise Development*, *23*(1), 163-188. https://doi.org/10.1108/JSBED-03-2014-0046

Antônio Rufino Júnior, C., Sanseverino, E. R., Gallo, P., Koch, D., Schweiger, H.-G., & Zanin, H. (2022). Blockchain review for battery supply chain monitoring and battery trading [Review Article]. *Renewable and Sustainable Energy Reviews*, *157*. https://doi.org/10.1016/j.rser.2022.112078

Cordova, R. S., Maata, R. L. R., Epoc, F. J., & Alshar'e, M. (2021). Challenges and Opportunities of Using Blockchain in Supply Chain Management [Article]. *Global Business & Management Research*, *13*(3), 204-217.

Deberdt, R., & Billon, P. L. (2021). Conflict minerals and battery materials supply chains: A mapping review of responsible sourcing initiatives [Review Article]. *The Extractive Industries and Society*, *8*(4). https://doi.org/10.1016/j.exis.2021.100935

Espinoza Pérez, A. T., Rossit, D. A., Tohmé, F., & Vásquez, Ó. C. (2022). Mass customized/personalized manufacturing in Industry 4.0 and blockchain: Research challenges, main problems, and the design of an information architecture [Article]. *Information Fusion*, *79*, 44-57. https://doi.org/10.1016/j.inffus.2021.09.021

Ethem, I., & Susanne, D. (2019). The Appropriation of Blockchain for Small and Medium-sized Enterprises [article]. *Journal of Innovation Management*, *7*(1), 26-45. https://doi.org/10.24840/2183-0606_007.001_0004

Exec. Order No. 13,817, 3 C.F.R. 1 (2017). https://www.federalregister.gov/d/2017-27899

Hashimy, L., Treiblmaier, H., & Jain, G. (2021). Distributed ledger technology as a catalyst for open innovation adoption among small and medium-sized enterprises [Article]. *Journal of High Technology Management Research*, *32*(1). https://doi.org/10.1016/j.hitech.2021.100405

Hildebrand, F., & Schneider, D. (2020). From Static to Continuous Supplier Qualification through Digital Technologies [Article]. *Von statischer zu kontinuierlicher Lieferantenqualifizierung durch digitale Technologie.*, *82*(7), 934-937.

Jiang, R., Kang, Y., Liu, Y., Liang, Z., Duan, Y., Sun, Y., & Liu, J. (2022). A trust transitivity model of small and medium-sized manufacturing enterprises under blockchain-based supply chain finance [Article]. *International Journal of Production Economics*, *247*. https://doi.org/10.1016/j.ijpe.2022.108469

Koyamparambath, A., Santillán-Saldivar, J., McLellan, B., & Sonnemann, G. (2022). Supply risk evolution of raw materials for batteries and fossil fuels for selected OECD countries (2000–2018) [Article]. *Resources Policy*, *75*. https://doi.org/10.1016/j.resourpol.2021.102465

Kumar Bhardwaj, A., Garg, A., & Gajpal, Y. (2021). Determinants of Blockchain Technology Adoption in Supply Chains by Small and Medium Enterprises (SMEs) in India [Article]. *Mathematical Problems in Engineering*, 1-14. https://doi.org/10.1155/2021/5537395

Lai, J.-Y., Wang, J., & Chiu, Y.-H. (2021). Evaluating blockchain technology for reducing supply chain risks [Original Paper]. *Information Systems and e-Business Management*, *19*(4), 1089-1111. https://doi.org/10.1007/s10257-021-00533-4

Li, B., Li, H., Dong, Z., Lu, Y., Liu, N., & Hao, X. (2021). The global copper material trade network and risk evaluation: A industry chain perspective [Article]. *Resources Policy*, *74*. https://doi.org/10.1016/j.resourpol.2021.102275

Liu, A., Liu, T., Mou, J., & Wang, R. (2020). A supplier evaluation model based on customer demand in blockchain tracing anti-counterfeiting platform project management

[Article]. *Journal of Management Science and Engineering*, *5*(3), 172-194. https://doi.org/10.1016/j.jmse.2020.06.001

Madhani, P. M. (2021). Enhancing Supply Chain Capabilities with Blockchain Deployment: An RBV Perspective [Article]. *IUP Journal of Business Strategy*, *18*(4), 7-31.

Morteza, G., & Mohammad, I. (2021). Digital transformation success under Industry 4.0: a strategic guideline for manufacturing SMEs [Journal]. *Journal of Manufacturing Technology Management*, *32*(8), 1533-1556. https://doi.org/10.1108/JMTM-11-2020-0455

Mugurusi, G., & Ahishakiye, E. (2022). Blockchain technology needs for sustainable mineral supply chains: A framework for responsible sourcing of Cobalt [Article]. *Procedia Computer Science*, *200*, 638-647. https://doi.org/10.1016/j.procs.2022.01.262

Nag, B., Han, C. and Yao, D.-q. (2014), "Mapping supply chain strategy: an industry analysis", Journal of Manufacturing Technology Management, Vol. 25 No. 3, pp. 351-370. https://doi-org.proxy-tu.researchport.umd.edu/10.1108/JMTM-06-2012-0062

Shao, L., & Jin, S. (2020). Resilience assessment of the lithium supply chain in China under impact of new energy vehicles and supply interruption [Article]. *Journal of Cleaner Production*, *252*. https://doi.org/10.1016/j.jclepro.2019.119624

Sunny, J., Undralla, N., & Madhusudanan Pillai, V. (2020). Supply chain transparency through blockchain-based traceability: An overview with demonstration [Article]. *Computers & Industrial Engineering*, *150*. https://doi.org/10.1016/j.cie.2020.106895

Teerasoponpong, S., & Sopadang, A. (2022). Decision support system for adaptive sourcing and inventory management in small- and medium-sized enterprises [Article]. *Robotics and Computer-Integrated Manufacturing*, *73*. https://doi.org/10.1016/j.rcim.2021.102226

Wang, K., Yan, F., Zhang, Y., Xiao, Y., & Gu, L. (2020). Supply Chain Financial Risk Evaluation of Small- and Medium-Sized Enterprises under Smart City [Article]. *Journal of Advanced Transportation*, 1-14. https://doi.org/10.1155/2020/8849356

Wu, Y., & Zhang, Y. (2022). An integrated framework for blockchain-enabled supply chain trust management towards smart manufacturing [Article]. *Advanced Engineering Informatics*, *51*. https://doi.org/10.1016/j.aei.2021.101522

Xu, Z., Zhang, J., Song, Z., Liu, Y., Li, J., & Zhou, J. (2021). A scheme for intelligent blockchain-based manufacturing industry supply chain management [Article]. *Computing*, *103*(8), 1771-1790. https://doi.org/10.1007/s00607-020-00880-z

Zheng, K., Zhang, Z., & Gauthier, J. (2022). Blockchain-based intelligent contract for factoring business in supply chains [Original Paper]. *Annals of Operations Research*, *308*(1-2), 777-797. https://doi.org/10.1007/s10479-020-03601-z

DECISION SCIENCES INSTITUTE

Literature Review of Benefits and Challenges of Learning Analytics in Higher Education

Ali Beheshti
University of Houston – Clear Lake
BeheshtiM3216@UHCL.edu

Xiaojun (Gene) Shan
University of Houston – Clear Lake
shan@uhcl.edu,

Ki Young Jeong
University of Houston – Clear Lake
JeongK@uhcl.edu

ABSTRACT

In this paper, we conduct a literature review of benefits and challenges of learning analytics (LA) in higher education. LA is an emerging field and the number of publications on this topic increases significantly over the past decade and remain a strong increasing trend. The benefits include efficient resource allocation, data-driven curriculum improvement, enhancement of students' learning, outcomes, behavior and processes, personalized learning, and improvement of instructors' performance. The challenges include data privacy and security issues, data collection.

KEYWORDS: Learning Analytics, Literature Review

INTRODUCTION

This paper aims to understand the causes and possible remedies to medical supply shortage by surveying a rich body of literature. In particular, this is a conceptual paper integrating the broad literature with a relatively unique perspective. Whereas there exist plenty of research studies focusing on particular types of medical products, few comprehensive reviews exist. Given that some policy regulations as mitigation strategies are suggested, which are likely to be broad affecting a range of medical products, it might be useful to compare the root causes and mitigation strategies for different medical products. The wide adoption of digital technologies in higher education is leading to the generation of a large amount of data. Educators and education researchers identify new opportunities to improve learning and its experiences by analyzing this data. An emerging field called Learning Analytics (LA) arise as a result. It could be defined as the usage of data, statistical analysis, and explanatory and predictive models to obtain insights about teaching and learning. Prakash (2015) states that LA involves analyzing data from learners and their activities to improve their learning experiences. Implementing LA allows higher education institutions to better understand their learners and the barriers to their learning. As a result, it could attempt to address those barriers to improve student outcomes and institutional success, which includes developing a larger and more diverse student population.

In addition, LA could provide teachers with better information on the quality of the educational content and associate learning activities with their teaching and assessment processes and outcomes, to enable continuous improvement. Niall (2016) mentions that obtaining better data on the student experiences could potentially enable institutions to identify and address issues such as inadequate feedback from the learners. Analytics could be used by teachers to monitor the performance and progresses of their students. They could then adapt their teaching if, for example, they had identified those students who are lagging behind the study. In case that the root cause is outside a single course, the higher-ed policymakers could use the analysis results to identify those large-scale problems and developing master plans that might control and reduce their impacts.

The rest of the paper is organized as follows: the next section provides the literature review as the main body of the paper. In particular, we first give a broad overview of research on LA, then discuss different benefits and challenges for LA. Finally, we provide conclusions.

LITERATURE REVIEW

Overview of Learning Analytics (LA)

Learning analytics is an emerging field, where advanced analytics tools and techniques are integrated to improve education, understanding and optimizing learning (Elias, 2011). As the quantity of data has increased, more researchers, educators, and businesses aim to develop new methods of data analysis. Manyika and Kiron (2011) mention that the recent emergence of Big Data and analytics is considered as having the potential to transform economies and improve organizational productivity and competitiveness.

According to Brown (2012), the process of systematically collecting and analyzing large data sets from online sources to improve learning processes is called learning analytics. LA is an emerging field in education and experts in online learning in American higher education predict that within the next few years learning analytics would be widely used in online education to identify students' patterns of behaviors and to improve students' learning and retention rates. Learning analytics would transform education, by changing the essence of teaching, learning, and assessment (Siemens and Long, 2011). Many vendors of learning technologies are providing analytics tools. For instance, Blackboard, Desire2Learn, Instructure and Tribal have all released analytics tools. There is also an active Moodle community focusing on learning analytics. The high-profile providers of Massively Open Online Courses (MOOCs) such as Coursera, Udacity and edX are all using analytics tools to support their operations.

There is also a growing research community around LA. An annual conference, Learning Analytics and Knowledge, has been organized (Long et al., 2011; Buckingham Shum, Gasevic, and Ferguson, 2012). Also, an international research network is set up: the Society for Learning Analytics Research (SOLAR). LA is defined by the SOLAR as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (SOLAR, p.1).

The rest of the paper is organized as follows: section 2 presents the methodology for this literature review. Section 3 provides the results based on the literature review. Section 4 draws some conclusions and provides some future research directions.

The following research questions guide this review:

1. What are the benefits of using learning analytics in higher education?
2. What are the challenges of using learning analytics in higher education?

METHODOLOGY

This study aims to investigate how LA has been used in higher education institutions and the outcomes obtained. Relevant research studies are collected from several databases such as Google Scholar, Springer Link, Science Direct, and Sage Pub using the key terms "academic analytics" and "learning analytics". After screening the resulted papers, inclusion and exclusion criteria are applied. In particular, the inclusion criteria to filter the papers are: 1) being peer-reviewed, 2) being published in English, 3) being related to learning analytics, and 4) being published after 2000 (learning analytics is emerging as an independent study area since approximately 2010). The exclusion criteria are: 1) being duplicates, 2) from grey literature such as news articles, and 3) academic and professional books. Then, the authors manually screen the title and abstract to exclude irrelevant papers. Since only one academic book is identified in Sage Pub, we exclude Sage Pub. Table 1 lists the number of articles found in each database.

Table 1. Number of articles found in Google Scholar, Springer Link, Science Direct and Sage Pub.

Database Name (Publication Year)	**Search Results for “academic analytics”**	**Search Results for “learning analytics”**
Google Scholar (2000-2022)	4,410	31,600
Springer Link (2000-2022)	37,242	1,417
Science Direct (2000-2022)	155	1,499
Sage Pub (2000-2022)	0	1 academic books

Trend of Learning Analytics

In the past decade, "learning analytics" has become one of the fields of educational technology that receives growing attention from educational researchers and practitioners (Hui and Kwok 2019). By searching for article-type studies from the SCOPUS database using the keyword “learning analytics,” it was found that the number of publications has increased at a fast pace since 2011 (see Fig. 1), showing the rapid growth of techniques, methods and applications of LA.

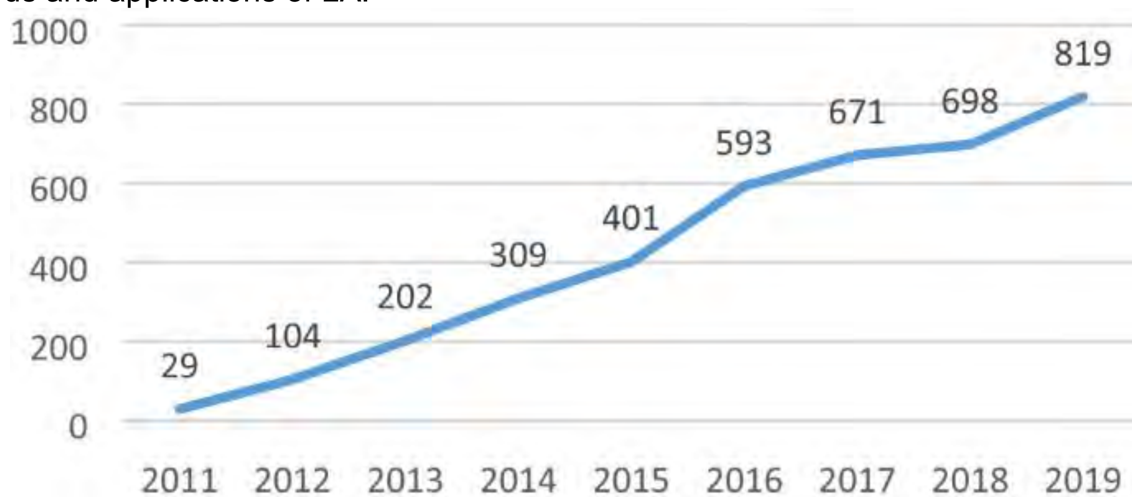


Figure 1. *Number of studies on learning analytics published in the SCOPUS database.*

The 2011 Horizon Report identified learning analytics as a possible key future trend in learning and teaching (Johnson, 2011). Learning analytics is not a genuine new research area. It actually, borrows from different related fields and synthesizes several existing techniques. According to figure 2 the number of studies on learning analytics and its tools shows a positive increasing from 2011 to 2019. For knowing more details about this trend and different tools and focuses that implemented by researcher, this time span divided to 10 brackets from 2011 to 2020.

Table 2. *Number of studies on learning analytics published in Google Scholar database.*

Learning Analytics Tools and Foci	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Data mining and big data		1*	3	3*	2*	3*	4*	2*	5*
Machine Learning				1	2	2*	3*	6*	7*
Learning analytics dashboards			1	1		2		1	4*
Natural Language Processing (NLP)					2*	1*	3*	2*	4*
E-learning[1]	1	2	1						
Visualization	3*				3*	2*	1*	3*	2*
Formative assessment	2*		1						
Learning design[2]	1	1*							1*
Massive Open Online Courses[3]							1	2	2*
Social Learning Analytics (SLA)		2							
Artificial intelligence	1*							1*	1
Data-driven decision making[4]		2*		1					
Technological innovation		1*	1		1				
Social network analysis (SNA)	1*								1
Students' behavior[5]	1	1*							
AR/VR tools, Video learning				1		1	2	1	3*
Mobile learning[6]			1						1
Adaptive learning[7]	1								

Learning Analytics Benefits in Education

Examination of the literature reveals how the use of big data is beneficial for higher education and includes various aspects from learning analytics that closely examine the educational process to improve learning. Another benefit includes the use of academic analytics that make alterations as a result of the application of algorithms to various points of data to improve learning. Through careful analysis of big data, researchers can determine useful information that can benefit educational institutions, students, instructors, and researchers in various ways. These stakeholder benefits include targeted course offerings, curriculum development, student learning outcomes and behavior, personalized learning, improved instructor performance, post-educational employment opportunities, and improved research in the field of education.

Identifying target courses

An initial benefit that evolves from using big data analytics in education is the ability of educational institutions to identify targeted courses that more closely align with student needs and preferences for their program of study. By examining trends in student enrollment and interests in various disciplines, institutions could allocate educational and teaching resources to programs that maximize student enrollment in the most needed areas of study. Schools can better predict graduate numbers for long-term planning of enrollment (Althubaiti & Alkhazim, 2014).

Curriculum improvement

Using big data allows instructors to make changes and adjustments to improve curriculum development in the educational system, such as in the use of curricular mapping of data (Armayor & Leonard, 2010). Through the analysis of big data, educators can determine weaknesses in student learning and comprehension to determine whether or not improvements to the curriculum may prove necessary. Instructors can engage in educational strategic planning to ensure that the learning curriculum targets student needs to maximize learning potential.

Student learning outcome, behavior, and process

Another key benefit of big data and text mining focuses on the ability of schools and instructors to determine student learning outcomes in the online learning as well as determine how to improve student performance (Bhardwaj & Pal, 2011). Researchers note that the use of educational data mining contributes to positive results in the learning process (AlShammari et al., 2013). Analysis of the data can help educators understand the student learning experience through learner interactions with technology tools such as E-learning and mobile learning (Hung & Zhang, 2012). Use of big data also reveals learning behavior, the impact on adaptive learning, and level of persistence (DiCerbo, 2014) in the learning process. By understanding the effects on learner outcomes, use of this data also reveals how to make improvements in student learning and performance in academic coursework. Therefore, LA allows instructors to evaluate forms of knowledge and adjust educational content accordingly.

Personalized learning

Arnold and Pistilli (2012) discussed an early intervention system that demonstrates the benefits and power of learning analytics. As an example, Course Signal provides students with real-time feedback. The components of students’ grades, demographic characteristics, academic background, and demonstrated effort are all addressed. The system employs a personalized email and a stoplight, specific color method to indicate progress or lack thereof. Using learning analytics, the concept of personalized learning reveals student success. Dietz-Uhler and Hurn (2013) asserted that course designers do not account for students who do not begin specific coursework at the same learning stage and who do not proceed, learn, and master course competencies at the same pace. Learning analytics allows faculty to use data collected by the learning management system to observe the frequency of student login. Instructors can also see student interaction within the course, total engagement, pace, and grades. These components serve as predictors of students' potential success or failure. Learning analytics allows for real-time reception of the pertinent data, review as well as the incorporation of data, and real-time feedback for every student.

Improved instructor performance

Using this data also helps to assess instructor performance (Mardikyan & Badur, 2011). The use of data provides an opportunity to improve instructor development so that instructors are better prepared to work with students in a technological learning environment. Through the acquisition of data generated from instructor usage of technology and research tools in online libraries (Xu & Recker, 2012), analysts can determine online behaviors by educators. Therefore, use of this information can help identify areas in need of improvement by the instructor to facilitate enhanced instructor-student interactions in the educational environment.

Learning Analytics Challenges in Education

The review of the literature revealed the LA challenges center on data tracking, data collection, data analysis, a connection with learning sciences, learning environment optimization, emerging technology, and ethical concerns regarding legal and privacy issues.

Data tracking

The digital tracking of information is a technique used by analysts to determine how best to present new learning opportunities as the wave of education continues to move forward into the second decade of the 21st Century. The tracking of big data represents the monitoring system. Current trend tracking indicators regarding the delivery and dissemination of instruction depend on the learning management system used by the institution. Platforms such as Moodle, Canvas, EPIC, and Blackboard have the capability to track the number of times an individual logs into the course room. These platforms also provide significant documentation to determine how involved the student was upon their login. Such tracking provides those who plan and implement new educational programs with valuable information. The monitoring reveals how engaging the curriculum presented is, as well as identifying areas that cause confusion (Brown, 2012).

Data collection

The collection of data can be a challenge when looking at LA. Nonetheless, it represents an important component in planning for continued implementation of educational program growth (Bottles et al. 2014). Educators must consider several elements. They must consider the availability of resources at a venue. Next, instructors must establish a viable social platform as it directly relates to interactions between learners to synthesize the educational content. Finally, instructors must discriminate whether the learner population possesses the requisite suitability for this type of learning environment and knowledge acquisition. Besides these challenges, gaps exist because of the inability to share proprietary information gathered by the institution. Further, another problem emerges because the creation of the ideal framework to disseminate educational curriculum takes teamwork, especially among the organizations bidding against one another to capture the learner population who want to engage in this type of learning experience.

Evaluation process

An important consideration of data collection concerns how learning analytics has become a force in the evaluation process. As greater amounts of educational resources become available online, there is a subsequent increase in the total data available regarding learning interactions. For learning analytics to help instructor evaluation to function appropriately, data needs to be delivered in a timely and accurate manner (Picciano, 2014). Learning analytics can provide powerfultools for developing meaning from interactions and actions within a higher education learning environment (Fournier, Kop, & Sitlia, 2011). With the unprecedented explosion of available data for online interactions, it is critical for the continued development of the evaluation process. LA can translate from other fields as interest in the data growth in education becomes more focused. Lias and Elias (2011) noted that statistical evaluation of rich data sources already exists within other professions and fields.

Data analysis

Technical challenges also exist from the assimilation of the data analysis because of the presentation format of the data. Erroneous data can skew the findings causing a misinterpretation of the overall population. Such scenarios are commonplace in the online learning environment. For example, an instructor may create a student profile to isolate an assignment that requires grading, test the ease of submission process, or to determine if there are any gaps in the presentation of the curriculum as it appears for students. Creation of a non-existent learner introduces redundant information that appears in the course without identification. This data does not represent student information but rather misinformation created by the instructor that flows into the big data pool of information (McNeely & Hahm, 2014). When manually conducting data analysis, this information can be easily identified from the population. However, working with data collection from the learning analysis vantage point adds a significant margin of error to the outcome of overall results.

Learning sciences connection

According to Pea (2014), personalized learning and learning opportunities demonstrate an inability to leverage learning analytics optimally; therefore, “the endgame is personalized Cyberlearning at scale for everyone on the planet for any knowledge domain” (p. 17). Ferguson (2012) asserted that to optimize and fully understand learning requires understanding how knowledge develops and how to support knowledge development. Further, researchers must understand the components of identity, reputation, and affect. Researchers must find ways to connect “cognition, metacognition, and pedagogy” (Vahdat et al., 2015, p. 299) to help improve learning processes. With a stronger connection to learning sciences, learning analytics can promote effective learning design.

Emerging technology

The full potential of learning analytics relating to learning requires continued and emerging technology that presently remains in the younger stages. This revelation presents a challenge as the technology continues to develop to stay constant with the growth of learning analytics.

Further, to fully understand the method and practice of teaching, more research is needed. Research focusing on learning analytics and pedagogy is still in the beginning stages (Dyckhoff et al., 2012).

Ethical and privacy issues

Another issue that emerges about learning analytics concerns the ethical, legal, and risk considerations (Kay et al., 2012). Because of dynamic changes in technology as well as how users store data and applications in cloud-based systems, "the challenges of privacy and control continue to affect adoption and deployment" (Johnson et al., 2011, p. 3). Further, the ethical and legal complexities of learning analytics challenge institutions that seek to implement their usage (Sclater, 2014a). For example, these considerations can include obvious areas of privacy considerations such as consent, data accuracy, how to respect privacy, maintaining anonymity, opting out of data gathering, and the potential effects to students. Additional concerns include data interpretation, data ownership, data preservation, sharing data with parties outside of the institution, and proper training of staff members regarding the handling of data (Sclater, 2014b). Further, the question becomes who owns this aggregate data, because having an infrastructure with the capacity to house large amounts of information becomes a daunting task (West, 2012). Because of these different issues, institutions must achieve a balanced approach to safeguard data while also assuring benefits to the educational process through the use of four guiding principles. These principles consist of clear communication, care, proper consent, and complaint (Kay et al., 2012). Institutions must demonstrate adherence to legal and ethical parameters to safeguard student privacy concerns while also achieving the educational goals for students and educators.

CONCLUSION

The study shows an increasing trend in the number of papers on LA in higher education. The results also show that most of these papers are published in journals related to engineering and technology, and they tend to use quantitative methodologies. In a context where technologies are present and mediate human behaviors across all spheres of life, the community of learning analytics in higher education is fast growing and is attracting attention and research efforts worldwide.

Critical studies of LA are needed so as to interrogate aspects such as data collection and analysis, and implications for students, teachers, managers, researchers and the academic community in general. Also, concerns and challenges identified in the paper invite us to revisit questions about the learning analytics in higher education; the marketization of education and the exploitation of data for business-like practices; accountability and audit processes that promote a conception of educational processes based on metrics; surveillance and the promotion of an Orwellian society in which students and teachers feel they are being surveilled; governance and management of data at institutional and national levels and their connection with educational policies. While it cannot be denied that technologies have created new environments for learning, through which students approach curriculum content and interact with others in a virtual way, the relationship between the extent to which the LA data is generated, gathered, and analyzed, and quality of learning remains unclear. Complex learning processes might be underplayed in the analytical techniques associated with LA so that the suggestion that LA should move from a technological focus towards a more educational focus (Ferguson 2012) deserves to be revisited.

In further examining issues on LA, aligned with previous literature (Daniel 2018), it is possible to venture a distinction between a practice-based community led by management units within higher education institutions and an academic community whose object of research study is LA. In other words, while managers and practitioners usually deal with learners' data in an everyday basis and develop strategies to improve student performance, prevent dropouts and predict completion rates, academics within the field of LA aim critically to examine both the technological tools mediating learning, the mathematical models, and the research methods in use so as to promote and develop learning theories. For both communities, LA have become a powerful tool to inform and improve learning through concrete interventions and actions.

Across both communities (practical and academic communities), there is a shortage of papers devoted to developing or expanding educational theories about students' learning (Ferguson and Clow 2017; Leitner et al. 2017; Viberg et al. 2018). This finding resurrects the issue as to the extent to which LA is about learning. Most of the empirical studies on LA seem focused on innovative methods of data collection, new ways of analyzing them, and the development of tools to support students' learning so exhibiting a rather pragmatic profile. The papers examined here within the thematic analysis are clear that educational and learning theories are insufficiently present in LA research. The role of educational theorists and critical approaches in understanding learning in its complexity are, therefore, crucial in overcoming this pending challenge.

Finally, an undue emphasis on metrics and quantification in research on LA legitimizes a technocratic perspective on learning that reinforces audit arrangements and a managerial discourse on learning in higher education. An active involvement of both teachers and students in contributing to design the learning environments and in assessing learning seem crucial. In other words, while practitioners, managers and academics are important stakeholders in the LA community, the presence of teachers and students needs to be secured and reinforced (Kollom et al. 2020). Also, coordination among institutional researchers, managers and academics is necessary so as to include a theoretical dimension. This will help in using pedagogy-based approaches and educational theories in understanding learning (rather than a data-driven approach only). In addition, a full development of LA would lie in a combination of quantitative and qualitative analyses (Al-Mahmood 2020). Qualitative studies could help in overcoming some of the main challenges that LA face such as the simplification of learning processes or the critique that LA is insufficiently sensitive to the time and place of the students' learning. Also, qualitative techniques might yield important insights in examining the teachers, students, managers and authorities' perceptions of LA, how students and teachers could be more actively involved, how privacy and confidentiality could be maintained, and how data could be better used to promote learning. There are limitations to be noted: 1) the number of analyzed papers is limited since it included only core collection and science indexes. This decision was made because our focus was on only learning analytics and its application in higher education, 2) the span of time for the search (2013–2019) is also limited, as we are interested in a trend in learning analytics in the past ten years, and 3) given these limitations and the analysis performed, the discussion and interpretations contained in the paper cannot be generalized to the entire LA community. A detailed qualitative analysis of different types of publications might provide a more comprehensive understanding of the extent to which LA investigates learning.

REFERENCES

Prakash, B., Hanumanthappa, D.M., & Kavitha, V. (2015). Big Data in Educational Data Mining and Learning Analytics. International Journal of Innovative Research in Computer and Communication Engineering, 2, 7515-7520.
Niall, S. (2016). Learning Analytics in Higher Education: A review of UK and international practice
Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute.
Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (2011). Analytics: The widening gap. MIT Sloan Management Review.
Elias T (2011) Learning analytics: definitions, processes and potential Ferguson R (2012) Learning analytics: drivers, developments and challenges. Int J Technol Enhanced Learning 4(5/6):304–317
Siemens, G. and Long, P. (2011). Penetrating the Fog. Retrieved online from http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education.
Long, Philip, George Siemens, Gra´inne Conole, and Dragan Gasˇevic´. 2011. Proceedings of the
1st International Conference on Learning Analytics and Knowledge (LAK11), Banff, AB Canada, Feb 27_Mar 01, 2011. New York: ACM.
Buckingham Shum, Simon, Dragan Gasevic, and Rebecca Ferguson. 2012. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver, BC, Canada, April 29May 02, 2012. New York: ACM.
Na, K. S., & Tasir, Z. (2017). Identifying at-risk students in online learning by analyzing learning behavior: A systematic review. In Big Data and Analytics (ICBDA), 2017 IEEE Conference, 118-123. IEEE.
Williams, P. (2014). Squaring the circle: a new alternative to alternative-assessment. Teaching in Higher Education, 19(5), 565-577.
Pascarella, E.T., Terenzini, P.T., 2005. How College Affects Students. Wiley & Sons, San Francisco, CA.
Astin, A.W., 1999. Student involvement: a developmental theory for higher education (reprinted from Journal of College Student Development, July, 1984). Journal of College Student Development 40 (5), 518e529.
Crosnoe, R., Johnson, M. K., & Elder, G. H. (2004). School size and the interpersonal side of education: An examination of race/ethnicity and organizational context. Social Science Quarterly, 85(5), 1259-1274.
Farooq, M. S., Chaudhry, A. H., Shafiq, M., & Berhanu, G. (2011). Factors affecting students' quality of academic performance: a case of secondary school level. *Journal of quality and technology management*, *7*(2), 1-14.
Cambridge University Reporter. (2003). Indicators of academic performance. Available: http://www.admin.cam.ac.uk/reporter/2002-03/weekly/5915/ [Accessed Feb. 12, 2014].
Fancsali, S. E., Zheng, G., Tan, Y., Ritter, S., Berman, S. R., & Galyardt, A. (2018). Using embedded formative assessment to predict state summative test scores. LAK '18 Proceedings of the 8th International Conference on Learning Analytics and Knowledge, 161-170.

Papamitsiou, Z., & Economides, A. A. (2016). Learning Analytics for Smart Learning Environments: A Meta-Analysis of Empirical Research Results from 2009 to 2015. Learning, Design, and Technology, 1-23.
Baharudin R and Luster T (1998). Factors Related to the Quality of the Home Environment and Children's Achievements. Journal of Family Issues.19(4):375-403.
Seyfried S F (1998). Academic Achievements of African American Preadolescents: The influence of Teachers Perceptions. American Journal of Community Psychology. 26(3):381-402.
Jeynes W H (2002). Examining the Effects of Parental Absence on The Academic Achievements of Adolescents: The Challenge of Controlling for Family Income. Journal of family and Economic Issues.23(2):132-141.
Hoschchild J L (2003). Social Class in Public Schools. Journal of Social Issues, 59(4):821-840.
Eamon M K (2005). Social- Demographic, School, Neighborhood and Parenting Influences on Academic Achievements of Latino Young Adolescents. Journal of Youth and Adolescence. 34(2):163-175.
McNeal Jr, R. B. (2001). Differential effects of parental involvement on cognitive and behavioral outcomes by socioeconomic status. The Journal of Socio-Economics, 30(2), 171-179.
Moore, D. S. (1990). Uncertainty. In Lynn Steen (Ed.), On the shoulders of giants: A new approach to numeracy (pp. 95–137). National Academy of Sciences.
Johnson, L., Levine, A., Smith, R., & Stone, S. (2010). The horizon report: 2010 edition (Report). Retrieved from http://www.nmc.org/pdf/2010-Horizon-Report.pdf
Baker, R. (2010). Data mining for education. International Encyclopedia of Education, 7, 112-118.
Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3–16.
Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40(6), 601-618.
Baker, R. (2010). Data mining for education. International Encyclopedia of Education, 7, 112-118.
Veyssieres, M.P. and Plant, R.E. Identification of vegetation state and transition domains in California's hardwood rangelands. University of California, 1998.
Jaschik, S. (2007). Texans and their tests. InsideHigherEd.com. Retrieved February 2, 2007. <http://www.insidehighered.com/news/2007/02/02/texas/>.
Lazar, S. (2015). The importance of educational technology in teaching. International Journal of Cognitive Research in Science, Engineering and Education, 3(1).
Eady, M., & Lockyer, L. (2013). Tools for learning: Technology and teaching. Learning to teach in the primary school, 71.
Saa, A. A. (2016). Educational data mining & students' performance prediction. International Journal of Advanced Computer Science and Applications, 7(5), 212-220.
F. Zhou, H.-L Duh, & M. Billinghurst, "Trends in augmented reality traching, interaction and display: A review of ten years in ISMAR,"in Mixed and Augmented Reality, ISMAR 7th IEE/ACM International Symposium, 2008, pp. 193-202
W. Matcha, & D. R. A. Rambli, "Preliminary Investigation on the Use of Augmented Reality in Collaborative Learning," in Informatics Engineering and Information Science, pp. 189-198, Springer Berlin Heidelberg, 2011.
M. Dunleavy, C. Dede, & R. Mitchell, "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning", Journal
Huang, R. (2019). The core value of artificial intelligence to promote the development of education. Digital Teaching in Primary and Secondary Schools, (8), 1.

Jackson, P. C. (2019). Introduction to artificial intelligence. Courier Dover Publications.
Wu, B. (2019). Personalized Education for Graduate Students Driven by Artificial Intelligence and Big Data. Teaching of Forestry Region, (7).
Johnson, L., Smith, R., Willis, H., Levine, A. & Haywood, K. (2011). The 2011 Horizon Report. Austin, Texas: The New Media Consortium.
Peters, M. A. (2019). Robotics in education and society.

DECISION SCIENCES INSTITUTE

Machine Learning Models for Predicting Age of Hens at Optimal Production of Eggs

Dinesh K. Sharma
University of Maryland Eastern Shore, USA
dksharma@umes.edu

Alok Kumar Shukla
D.A.V. College, Kanpur, India
alokshukladav@gmail.com

Saif Ali Khan
Babasaheb Bhimrao Ambedkar University, Lucknow, India
saif.rs.stats@email.bbau.ac.in

S. K. Yadav[1]
Babasaheb Bhimrao Ambedkar University, India
Email: drskystats@gmail.com
[1]Corresponding Author

ABSTRACT

In this paper, we suggest a mixed linear regression model by using two linear regression models along with three Machine Learning (ML) models for the analysis and prediction of the age of hens at optimum production of eggs. The parameters of the mixed linear regression model have been estimated using the ordinary least square method. Using the different components of the two linear regression models, we have formed different models, and the best models have been obtained based on the values of different fitting measures, such as the coefficient of determination and the residual sum of squares. Different regression and machine learning fitted models are verified on real egg production data, and the best model is recommended for practical utility.

KEYWORDS: Linear regression model, Machine learning models, Least square method, Fitting measures, Prediction.

INTRODUCTION

The fastest-growing sector of India's agricultural industry, poultry farming, is crucial to the nation's economy. Numerous genes work together with biochemical, anatomical, and physiological processes to produce eggs. Various factors, including breed or strain, age of the hens, photo-refractoriness, broodiness, moulting, diet, and other environmental factors, influence the number of eggs laid by hens (Okereke and Nwogu, 2010). To determine the capacity for egg production, the daily egg production is summed on a weekly, bimonthly, or monthly basis because the hens do not lay eggs every day and take breaks. The egg production curve against time climbs quickly to a maximum when everyday egg production is summed up on such a time scale. After that, it falls to a minimum amount at the end of the year. At a given age, the quantity of eggs normally peaks before declining at the end of the laying season. Utilizing part data to predict egg production age early in the laying cycle enables poultry breeders to choose breeding hens early, which lowers the price of egg production and day-old chicks. Numerous published statistical models explain egg production in laying chickens. Cason (1990, 1991) studied egg production forecasting.

Acha (2010) emphasized the significance of improving bird production. Ahmad (2011) worked on egg production forecasting. Narinc *et al.* (2014) emphasized egg production curve analyses in poultry science. Yakubu *et al.* (2018) focused on modeling egg production using classification regression tree approaches and linear, quadratic, and artificial neural networks. Sharifi *et al.* (2022) conducted a study to evaluate eight mathematical models for egg production and egg weight. Many machine-learning algorithms have also been applied in the literature for the best egg production. An artificial neural network was utilized by You *et al.* (2021) to forecast the likelihood of oviposition events in broiler breeder hens that were precision fed, and Machine Learning (ML) was employed by Magemo *et al.* (2022) to predict the production of eggs in chicken farms.

In this paper, a new regression model has been introduced by combining two existing linear regression models, and three machine learning models have also been used for the optimal production of eggs based on the age of hens. The properties of the models are studied, and the best-fitted models are obtained, which may be used to predict the optimal production of eggs as per the age of the hens. The whole paper has been presented in different sections.

LITERATURE REVIEW

Several researchers used many regression models as well as machine learning (ML) models for the analysis and prediction of the age of hens at optimal egg production. Some of the research articles dealing with the different regression and ML models for the optimal production of eggs are presented in Table 1.

Table 1: List of Research Contributions in the Field of Regression and Machine Learning Methods

YEAR	REFERENCES	JOURNAL
1971	Gavora *et al.*	Poultry Science
1980	Adams and Bell	Poultry Science
1982	Gavora *et al.*	British Poultry Science
1997	Fialho and Ledur	British Poultry Science
2000	Grossman *et al.*	Poultry Science
2008	Ahmadi *et al.*	Journal of Animal and Veterinary Advances
2009	Ahmad, H. A.	Journal of Applied Poultry Research
2011	Ahmad, H. A.	Journal of Applied Poultry Research
2015	Felipe *et al.*	Poultry Science
2015	Chaudhari and Tingre	Indian Journal of Animal Research
2018	Yakubu *et al.*	Livestock Research for Rural Development
2019	Guo *et al.*	Animal Reproduction Science
2020	Akilli *et al.*	Tropical Animal Health and Production
2020	You *et al.*	Computers and Electronics in Agriculture
2020	Omomule *et al.*	Computers and Electronics in Agriculture
2021	Nyalala *et al.*	Poultry Science
2023	Welch *et al.*	Animals
2023	Yang *et al.*	Artificial Intelligence in Agriculture
2023	Bumanis *et al.*	Procedia Computer Science

GOODNESS OF FIT OF DIFFERENT MODELS

The following fitting measures are used to assess the goodness of fit of several models, and the best-fit regression model is then determined based on the values of the fitting metrics. Different fitting measures, along with their formulae, are presented in Table 2.

Table 2: Measures of goodness of fit with their formulae

S. No.	Name of Fitting Measure	Formula
1.	Coefficient of Determination	$R^2 = 100\left(1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}\right)\%$ Where, y_i is observed value of $\hat{y}_i$ is the fitted value by the model
2.	Adjusted Coefficient of Determination	$R^2_{adj} = 100\left(1-\left(\frac{n-1}{n-p}\right)\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}\right)\%$ Where, n is the number of observations and p is the number of parameters of the model.
3.	Mean Squared Error (MSE)	$MSE = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-p}$

ANALYSIS OF EXISTING MODELS

In this section, we have applied various discussed existing regression models given by Pandey and Kumar (2014) and Nwogu and Acha (2014) to the data set of eggs in hand, given by Nwogu and Acha (2014). The values of the fitting measures R^2 and the Mean Squared Error (MSE) for these regression models are presented in Table 3.

Table 3: Values of R^2 and MSE for Different Regression Models

Model	R^2	MSE
Pandey and Kumar (2014)	91.4017	0.9979
Pandey and Kumar (2014)	91.1469	1.0275
Nwogu and Acha (2014)	87.5000	1.4480

SUGGESTED REGRESSION AND MACHINE LEARNING MODELS

Motivated by the existing models in the literature, we suggest the following linear regression model for the analysis and prediction of age at optimal production of eggs as:

$$Y_i = A + B\,X_i + C\frac{1}{X_i} \tag{1}$$

Where, A is the intercept term, B and C are the regression coefficients. This model has also been used by us in Misra *et al.* (2009) in sampling theory.

Table 4 below lists the estimated parameters and fitting metrics for the proposed model using the ordinary least square approach.

Table 4: Parameter Estimates and Fitting Measures of the Suggested Model

Parameter	Estimate
A	42.9661
B	-0.3842
C	-622.405
R^2	91.1422
MSE	0.9959

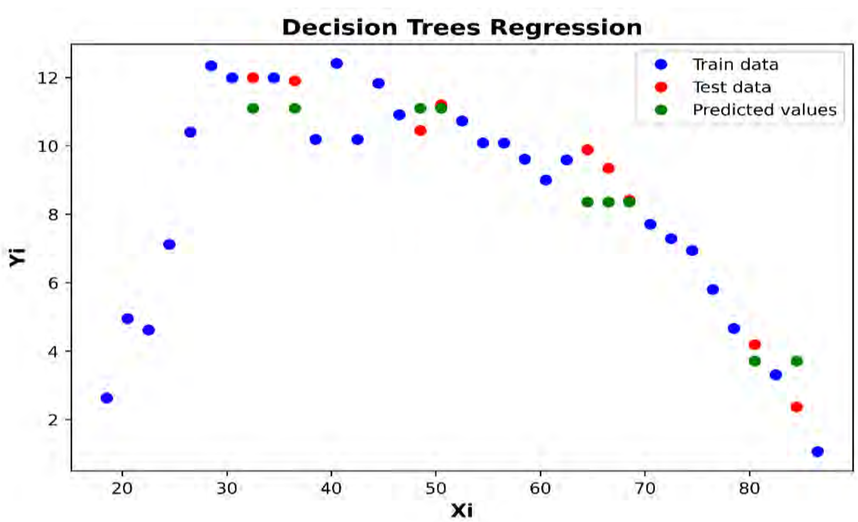
The MSE of the suggested model is the least among the competing models of Pandey and Kumar (2014) and Nwogu and Acha (2014). Therefore, the suggested model is best among the competing models in Table 3. In addition, the Linear Regression model, Linear Regression model with Polynomial Features (degree = 2) for Polynomial Regression model, and Decision Trees Machine Learning (ML) models were employed to forecast the age of hens at optimal egg production.

Table 5 displays the MSE values for all three machine learning models, reflecting the accuracy of their predictions for the optimal egg production age of hens. The MSE ranges over several runs are also presented, illustrating the variety in these models' performance when the program is executed repeatedly.

Table 5: Fitting Measures of Machine Learning Models

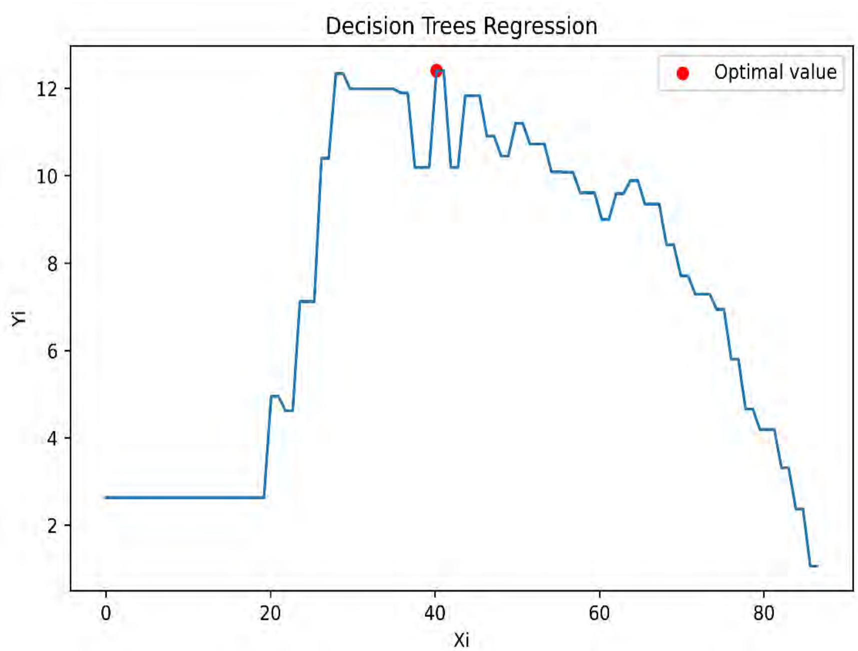
Models	MSE	Optimum Age	MSE Range
Linear Model	7.124	-	(5.0-20.0)
Polynomial Features	1.2455	48.06	(1.2 - 6.5)
Decision Trees	0.7992	40.19	(0.5 - 4.5)

Figure 1: Prediction of age of hens using the Decision Trees Model



Furthermore, the figures of Linear Regression model, Linear Regression model with Polynomial Features (degree=2) and Decision Trees ML models have also been presented for predicting the age of hens at the optimal egg production. Among the machine learning models, the Decision Trees is the best as it has least MSE. Further Decision Trees is also the best among the regression models as well as the machine learning models. Thus the prediction by the Decision Trees model will be closer to the reality. Figure 1 represents the graphs of the fitting of Decision Trees model showing fitting of trained data, test data and predicted values while Figure 2 represents the predicted graph along with the optimal value.

Figure 2: Prediction of age of hens at optimal production of eggs using Decision Trees model



RESULTS AND DISCUSSION

From the results (Table 3), it is observed that the values of R^2 and MSE for Pandey and Kumar (2014) regression model (2) are 91.4017 and 0.9979, respectively, and for Pandey and Kumar (2014) regression model (3) are 91.1469 and 1.0275, respectively, while these for Nwogu and Acha (2014) regression model (4) are 87.50 and 1.4480, respectively. The values of R^2 and MSE from Table 4 for the suggested regression model (4) are 91.1422 and 0.9959 respectively. From Table 5, it may be observed that the MSE and the range of MSE are 7.124 and (5.0-20.0) respectively. The MSE, Optimum Age and range of MSE for Polynomial Features are 1.2455, 48.06, (1.2 - 6.5) while for Decision Trees are 0.7992, 40.19, (0.5 - 4.5) respectively.

CONCLUSION

In this study, we have fitted different Regression and Machine Learning models for the prediction of ages of hens at optimal production of eggs. The fitting measures like Coefficient of determination R^2 and MSE are calculated for the competing regression models of Pandey and Kumar (2014) and Nwogu and Acha (2014) along with the suggested regression model and the machine learning models. As per the results and discussion, it is evident that the suggested regression model is best among the competing regression models of Pandey and Kumar (2014) and Nwogu and Acha (2014). Further, it is observed that the Decision Trees Machine Learning model is better than the suggested regression model as it has least MSE among the class of competing and suggested regression models. Thus, the Decision Trees Machine Learning model closely predicts the age of hens at optimal egg production, which may be utilized for best policy making.

REFERENCES

Acha, C.K. (2010). Impact of Housing and Age of Birds on Egg production, *International Journal of Numerical Mathematics*, 5(2), 275-298.

Adams, C. J. and Bell, D. D. (1980). Predicting Poultry Egg Production. Poultry Science, 59(4), 937–938. https://doi.org/10.3382/ps.0590937

Ahmad, H. A. (2009). Poultry growth modeling using neural networks and simulated data. Journal of Applied Poultry Research, 18(3), 440–446. https://doi.org/10.3382/japr.2008-00064

Ahmad, H. A. (2011). Egg Production Forecasting: Determining Efficient Modeling Approaches, *Journal of Applied Poultry Research*, 20(4), 463–473. https://doi.org/10.3382/japr.2010-00266

Ahmadi, H. and Golian, A. (2008). Neural network model for egg production curve. Journal of Animal and Veterinary Advances, 7(9).

Akilli, A. and Gorgulu, O. (2020). Comparative assessments of multivariate nonlinear fuzzy regression techniques for egg production curve. Tropical Animal Health and Production, 52(4), 2119–2127. https://doi.org/10.1007/s11250-020-02226-5

Bianco, V., Manca, O., & Nardini, S. (2009). Electricity consumption forecasting in Italy using linear regression models. Energy, 34(9), 1413–1421. https://doi.org/10.1016/j.energy.2009.06.034

Bumanis, N., Kviesis, A., Tjukova, A., Arhipova, I., Paura, L. and Vitols, G. (2023). Smart Poultry Management Platform with Egg Production Forecast Capabilities. Procedia Computer Science, 217, 339–347. https://doi.org/10.1016/j.procs.2022.12.229

(A complete list of references is available upon request.)

DECISION SCIENCES INSTITUTE

Machine Learning Using Spreadsheets: K-Means Cluster

Leong, Thin Yin
Singapore University of Social Sciences
Email: tyleong@suss.edu.sg

Ma, Nang Laik
Singapore University of Social Sciences
Email: nlma@suss.edu.sg

ABSTRACT

Machine learning as a phenomenon has gone viral in recent years. However, the tools offered remain highly technical and not accessible to those without rigorous computer science or business analytics training. With simplicity and accessibility in mind, this paper renders the clustering method, a fundamental machine learning methodology, on the ubiquitous and easily comprehensible spreadsheet without macros or add-ins like Solver. It can be achieved using the elegant though obscure recursive computation feature. The paper can be used to form the core content for introducing machine learning to non-technical audiences, particularly those in Business and the Social Sciences.

KEYWORDS: machine learning, K-means cluster, business analytics, spreadsheets

INTRODUCTION

Machine Learning has been getting lots of attention recently in academia and the public media and businesses. Machine Learning can be broadly categorised into three types, namely supervised Learning, Unsupervised Learning, and reinforcement learning. Supervised learning uses the training data to learn the relationship between input variables (X) and output variables (Y). In this case, the output variable is labeled. If the output variable is a continuous real value, such as predicting customer spending or the student's score, we use prediction or classification methods.

On the other hand, if the outcome of a given sample is categorical such as "Approve" or "Reject," we will use classification methods. Unsupervised learning is to explore the data and draw inferences from the dataset without the target or response variable or label. The data only has the input variables without the corresponding output value. Unsupervised learning is helpful if we want to uncover the unknown relationship between the underlying structures of the data. Cluster analysis is the most common unsupervised learning. It is used to explore the data and group the data points into several groups (called clusters) so that the data in the same group are more similar but very different from the data in different groups.

Finally, Reinforcement Learning (RL) allows software agents to take the next best actions in an environment based on its current state to maximize some cumulative reward. RL usually learns the optimal actions through trial and error or random movement. The labeled input/output pairs may not be presented like supervised learning in RL, and sub-optimal actions need not be explicitly corrected.

Data mining methods are not new; their resurgence is primarily due to the broad-scale, high-speed, and low-cost computation and wireless communications and the massive availability of live data from the Internet and smartphone usage. As a result, billions of dollars of investments by vendors and users alike are going into its development and application.

The original Coursera course by Ng (2017) is massively popular, with 1.7 million enrollments since the course began in 2012. Many more people continue to take it online or view it via YouTube to learn the basic concepts and techniques (supervised and unsupervised learning, neural network, support vector machines and clustering). Other topics include statistical techniques (single and multiple linear and logistic regressions) and mathematics (linear algebra, forward propagation and backpropagation, and gradient descent optimization.

As the onerous list of machine learning topics above suggests, this is a highly technical course for scientists and engineers. However, learners are already spared from R or Python programming and associated TensorFlow and other libraries. The inherent knowledge, skills, and tools offered by software vendors remain highly technical and not accessible to people without training in Computer Science or Business Analytics. Machine learning would only become broadly used if the average business executive, non-technical undergraduate, or even high-school student could understand it—no complex mathematics or vague sales pitch. Perhaps it might be helpful if learners could simply interact with machine learning, figuratively from scratch, building simple models and not just blind application.

With the better acceptance of this methodology as an acquired skill, many more may want to spend the intensive effort to learn the required challenging mathematics and cryptic programming. After trial application explorations, organizations are willing to put aside a substantial budget to employ skilled people to apply them to challenging settings. As the history of AI applications informs us, particularly in medicine, unless end-users can understand, accept and trust it, all "better than human experts" abilities, however, well-proven, verified, and validated, would come to naught.

With simplicity and accessibility in mind, this paper shows how the K-means cluster, a fundamental machine learning methodology, can be rendered on ubiquitous and more comprehensible off-the-shelf spreadsheet applications directly without macros or add-ins, or Solver. We will demonstrate the application of the K-means cluster using spreadsheets on a familiar problem: from Winston & Albring (2019), which contains 3000 transactions' data of 250 customers. Each row (Transaction) records the transaction of sale ($) of the five footwear categories (Type) by customer (CusID). The goal is to cluster customers by the similarity of their buying behavior.

Our paper's key contribution is to show non-technical end-users more comprehensively what the K-means cluster is and how to do it simply in Spreadsheets. It refreshes upon the efforts of others to promote the use of spreadsheets as a tool for learning and applying basic K-means clustering concepts. In particular, the paper's novelty is our use of the elegant though obscure, recursive computation spreadsheet feature and so do away with the need for Solver. Solver's use for the K-means cluster is still comprehensively explained and demonstrated as it would be handy to those who cannot fully understand or accept recursive computation.

The rest of the paper is organized as follows: Section 2 contrasts our approach against the related effort by others, and Section 3 explains the base model. Next, Section 4 shows the complete model by varying the value of k from 2 to 10 and summarizes the results. Finally, the paper ends with a broad discussion on research motivation and future work, to introduce machine learning to non-technical students, particularly those in Business and the Social Sciences.

LITERATURE REVIEW

The reasons for using spreadsheets are many (Leong & Cheong 2008). Briefly, a spreadsheet permits people with simple mathematical skills and no programming abilities to perform a detailed modeling of business challenges by making algebraic, statistical, and probabilistic computations appear arithmetic-like. Its interface is simple and highly interactive, and its ubiquitous presence in offices led many analysts to consider it the tool of choice for exploring business opportunities. The authors use spreadsheets to provide viable means for the less mathematically- and technically inclined to explore realistic business scenarios and resolve them without the use of add-ins and macro programming. Problems addressed to date include multi-server queues (Leong 2007a), and data resampling, and Monte-Carlo simulation (Leong 2007b, Leong & Lee 2008).

Nowadays, many university professors are adapting spreadsheets as their primary computing platform to support the teaching of mathematics, statistics, and management science courses. Graduates can also apply spreadsheets easily for quick data analysis and understanding complex solution approaches. Moreover, even people without formal training but only direct experience with spreadsheets can pick up advanced modeling techniques (Leong & Cheong 2008, 2015). Therefore, the critical advantage of using spreadsheets is that it would allow a wider end-user audience to experience what machine learning is and how applications may be easily constructed and flexibly adapted for use.

However, there appear to be a few academic papers on using spreadsheets in machine learning. In addition, there are some valuable articles by professionals on web forums and blogs that address the use of spreadsheets for machine learning. However, they cover only basic statistical analysis approaches such as Linear Regression and Logistic Regression. For example, Granville (2017) uses complex array functions (i.e., CORREL, COVAR, and LINEST). Unfortunately, this approach makes the models opaque, as users cannot see how detailed computations are done. Still, others like Roberts (2018) introduced the use of the PyXLL Add-in that embeds Python in Excel to extend the spreadsheet's functionality to take advantage of the Python ecosystem.

MODEL DEVELOPMENT

Basic Model, k = 4

The Forgy method partitions data points into K clusters, randomly assigning each point to a cluster. Then, it computes the centroid of each cluster and the cluster of the nearest centroid. This process is repeated until reassignments cannot be made.

This heuristic is quite a challenge to do it a spreadsheet wisely resorted to using Solver. K randomly selected points may be assigned as centroids. The Solver then changes the assignment labels (all restricted to whole numbers) to find an assignment with the lowest average distance to assigned clusters' centroids. Additional constraints may be included to ensure that the capacity of clusters is not violated.

The workbook examines how a spreadsheet approach to do clustering without Macro, add-in or Solver. What if cluster capacity needs to be included?

The case examined from Winston & Albright (2019), contains the 3000 transactions' data of 250 customers, with each transaction recording the sales from a customer on the five footwear categories. The goal is to cluster customers by the similarity of their buying behaviour.

Here is the sample of the dataset. The first column is Transaction (transaction number), followed by CustID (Customer ID), Type (of shoes purchased, namely, Athletic, Casual, Dress, Sandal and Work), Spent (total amount spent on the transaction). The purpose is to cluster the customers and make sense of the resulting clusters.

Table 1: Total transaction shown in "Data" worksheet.			
Transaction	**CustID**	**Type**	**Spent**
1	210	Sandal	29
2	7	Work	74
3	220	Dress	134
4	93	Athletic	150
5	66	Athletic	168
6	232	Dress	125
7	132	Work	72
8	82	Casual	102

Firstly, we need to consolidate the transaction data by CustID. We use an Excel function called SUMIFS. SUMIFS(sum_range, criteria_range1, criteria1, criteria_range2, criteria2)

In this case, the sum range is the Spent column, where criteria_range1 is CustID range which is equal to CustID specified and another criteria_range2 is the Type of shoe range which is equal to Shoe type of Row14.

The formula in Cell C15 is as followed:

=SUMIFS(Data!$D$2:$D$3001,Data!$B$2:$B$3001,$B15,Data!$C$2:$C$3001,C$14)

We need to lock the data in sum_range as an absolute reference, criteria_range1, and criteria_range2. For criteria1, we need to lock column B as $B15, and criteria2 is at Row 14, so we need to use mixed reference as C$14. After we have fixed the cell correctly, we can fill up the whole table for 250 customers, and the result will be shown in total $ sale for each customer shown in Figure 1.

As the Spent amount varies in magnitudes by the shoe type, we need to standardize it to provide better data for clustering.

Figure 1: To consolidate data in the "Data" worksheet by customer ID in the "Data(2)" worksheet

Shoe Sales

Comments

1. The transaction data consolidated by customers below.
2. As the data values vary in magnitudes by the shoe categories, they need to be standardised to provide better data for clustering.

Min	0	0	0	0	0
Max	2267	713	3206	601	2018
Avg	302	213	526	120	275

Sales $ Total

CustID	Athletic	Casual	Dress	Sandal	Work
1	1257	77	719	42	184
2	0	499	190	160	170
3	0	42	1768	0	0
4	1862	381	0	0	0

We show the min, max, and average for each shoe type, such as Athletic, Casual, Dress, Sandal, and Work. For example, the average for Dress shoes is $526, which is more than four times of Sandal, and Athletic shoes are the second-highest in the Spent ($) amount.

Next, we want to develop the K-means cluster using the template given in Figure 2. Currently, the user input K value is 4 in Cell D4, and users can choose from the list between 2-10 clusters.

Figure 2: Template for the K-means clusters in the "Proto" worksheet

Shoe Sales

Mean Dist	K	Iteration Start
	4	No

	Cluster	1	2	3	4	5	6	7	8	9	10
X1	Athletic										
X2	Casual										
X3	Dress										
X4	Sandal										
X5	Work										

	Values in [0, 1]					Assigned		Next	Distance to										
CustID	Athletic	Casual	Dress	Sandal	Work	Cluster	Dist	Cluster	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
1																			
2																			
3																			
4																			
5																			
6																			

We need to standardize the Spent on each shoe Type between 0 and 1. The formula we use is $s_{ij} = \frac{a_{ij}}{\max(a_{ij}), for\ each\ j}$, where a_{ij} is the spent by each customer i (i=1,..,250) on Type j (j=1,2,3,4,5) and the denominator is the maximum spent on category j, we will get the result between 0 and 1.

The formula in Cell C15 is: ='Data(2)'!C15/'Data(2)'!C$9

C$9 is locked as it is the maximum spent on Athletic, and we should display the value in 2 decimal places. We can fill up the table from Range C15:G264. We will see a similar result as shown in .

Figure 3: Standardised Spent between 0 and 1

We set the iteration to NO first. Then, we need to set up a formula to assign the cluster, compute the centroid and distance from the centroid.

If the iteration is "No", the cluster assigned is empty. Otherwise, it is the same as the next cluster. Cluster assignment in Cell H15 is =IF(H4="No","",J15)

Table 2: Initial cluster assignment is empty for all						
CustID	Athletic	Casual	Dress	Sandal	Work	Cluster
1	0.55	0.11	0.22	0.07	0.09	
2	0.00	0.70	0.06	0.27	0.08	
3	0.00	0.06	0.55	0.00	0.00	
4	0.82	0.53	0.00	0.00	0.00	
5	0.54	0.22	0.00	0.14	0.07	
6	0.29	0.21	0.50	0.53	0.00	

Next, we will compute the centroid for each cluster. Since the initial assignment of the cluster is empty, we need to set some initial centroid values arbitrarily.

Figure 4: To set up some initial cluster centroid using Span

	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM
1			Initial/Default Cluster Coordinates									
2	Span		1	2	3	4	5	6	7	8	9	10
3	0.70	x1	0.15	0.43	0.80	0.53	0.74	0.52	0.47	0.57	0.21	0.29
4	< Max Coordinate	x2	0.38	0.31	0.22	0.18	0.82	0.21	0.41	0.26	0.55	0.51
5		x3	0.44	0.82	0.36	0.21	0.39	0.59	0.26	0.26	0.19	0.64
6		x4	0.35	0.29	0.22	0.72	0.71	0.69	0.81	0.18	0.62	0.23
7		x5	0.73	0.21	0.29	0.77	0.35	0.20	0.41	0.21	0.73	0.40

Span in Cell AB3 is the spread, which is less than the maximum coordinate of 1 in this case, and we use the formula to get the initial value.

The formula in Cell AD3 is = =0.5-$AB$3/2+RAND()*$AB$3

After fixing the $AB$3, we can drag it to fill up the whole table.

The centroid of a cluster k, c_k is computed as = $\frac{\sum_{i=1}^{n}\sum_{j=1}^{m} s_{ij}}{n}, \forall(i,j) \in k$ where s_{ij} is the standardize number of Spent on category j by customer i, which belongs to the cluster k.

Table 3: Sample data to compute the centroid						
CustID	Athletic	Casual	Dress	Sandal	Work	Cluster
1	0.55	0.11	0.22	0.07	0.09	1.0
5	0.54	0.22	0.00	0.14	0.07	1.0
9	0.15	0.39	0.19	0.12	0.32	1.0

Assuming, we have these three customers in cluster 1, centroid for cluster 1 based on the information above is:

=(0.55+0.54+0.15/3, 0.11+0.22+0.39/3, 0.22+0+0.19/3, 0.07+0.14+0.12/3, 0.09+0.07+0.32/3)
= (0.41, 0.24, 0.14, 0.11, 0.16)

In Excel, we will use the AVERAGEIFS function to compute the centroid for all cluster 1 in Range K4:K8. We can fill up the table on top using the same formula to find the centroid for different clusters. We will see that some of the centroids have an error as the denominator is Zero. To avoid any error, we use IFERROR and replace it with a random number between 0 and 1.

Formula in K4 (Athletic),= IFERROR(AVERAGEIFS($C$15:$C$264,$H$15:$H$264,K$3), AD3)
Formula in K5 (Casual), =IFERROR(AVERAGEIFS($D$15:$D$264,$H$15:$H$264,K$3), AD4)
Formula in K6, (Dress),=IFERROR(AVERAGEIFS($E$15:$E$264,$H$15:$H$264,K$3), AD5)
Formula in K7, (Sandal), = IFERROR(AVERAGEIFS($F$15:$F$264,$H$15:$H$264,K$3), AD6)
Formula in K8, (Work), =IFERROR(AVERAGEIFS($G$15:$G$264,$H$15:$H$264,K$3), AD7)

After we have computed the centroid, we want to check if the distance from the point to the cluster's centroid is the smallest. If it is not the smallest, it will be assigned to the clusters with minimum distance.

Distance from the centroid of cluster = $\sqrt{\sum_{j=1}^{m}\left(s_{ij}-c_k\right)^2}$ where c_k is the centroid of the cluster k and s_{ij} is the standardize number of Spent on category j by customer i, which belongs to the cluster k.

The formula in Cell C15, distance to K15 is as follows:
=SQRT(($C15-K$4)^2+($D15-K$5)^2+($E15-K$6)^2+($F15-K$7)^2+($G15-K$8)^2)

The formula in Cell C15, distance to L15 is as follows:
= SQRT(($C15-L$4)^2+($D15-L$5)^2+($E15-L$6)^2+($F15-L$7)^2+($G15-L$8)^2)

The formula in Cell C15, distance to M15 is as follows:
= SQRT(($C15-M$4)^2+($D15-M$5)^2+($E15-M$6)^2+($F15-M$7)^2+($G15-M$8)^2)

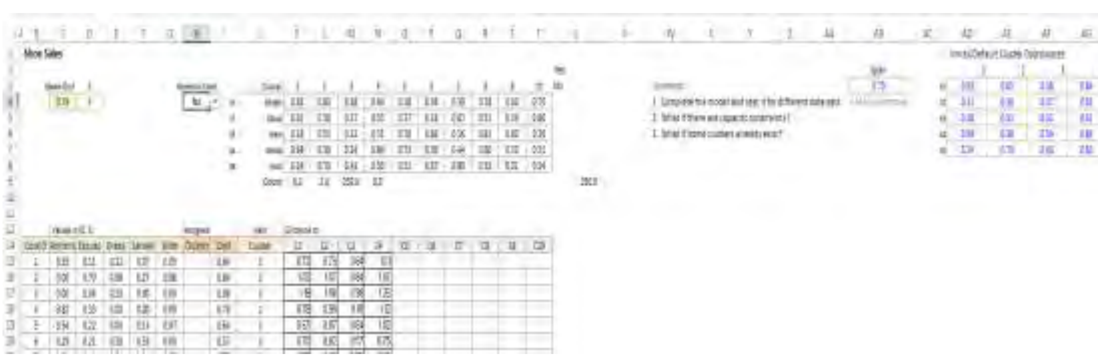
The formula in Cell C15, distance to N15 is as follows:
=SQRT(($C15-N$4)^2+($D15-N$5)^2+($E15-N$6)^2+($F15-N$7)^2+($G15-N$8)^2)

After we computed the distance, we wanted to find the minimum distance and assign the customer to this cluster minimum distance.

The formula to compute the distance of the customer from the centroid in Cell I15 is, = =min(K15:N15,,H15)
The formula for Next Cluster assigned, in Cell J15 is, =MATCH (I15,K15:N15,0)

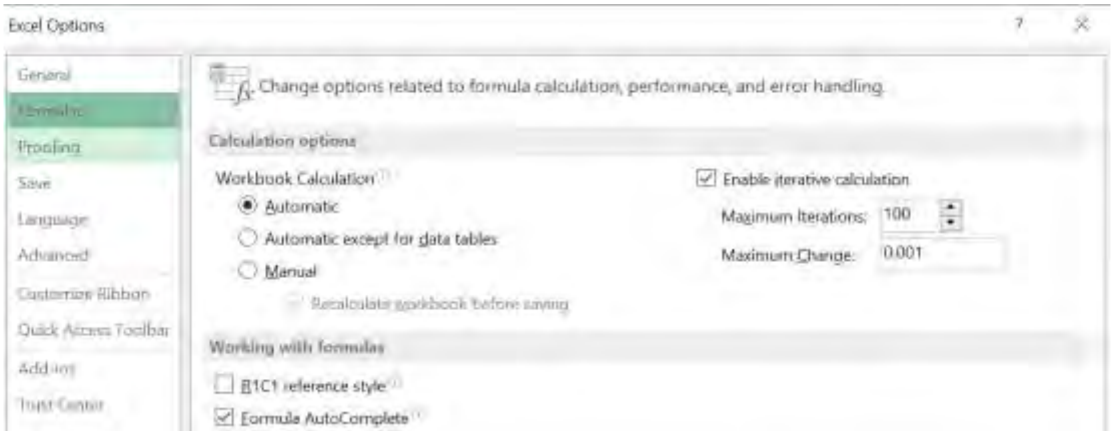
We can then fill up the whole table. We want also to find out the number of customers in each cluster, using COUNTIF function in Cell K9, = =COUNTIF($H$15:$H$264,K$3)

Figure 5: After 1st iteration with initial cluster coordinate



We now make use of the Excel iterative calculation. Go to Excel→ File→ Options → Formulas, Check on "Enable iterative calculation," and set to 100 and maximum change to 0.001 as shown in Figure 6.

Figure 6: Excel options for iterative calculation

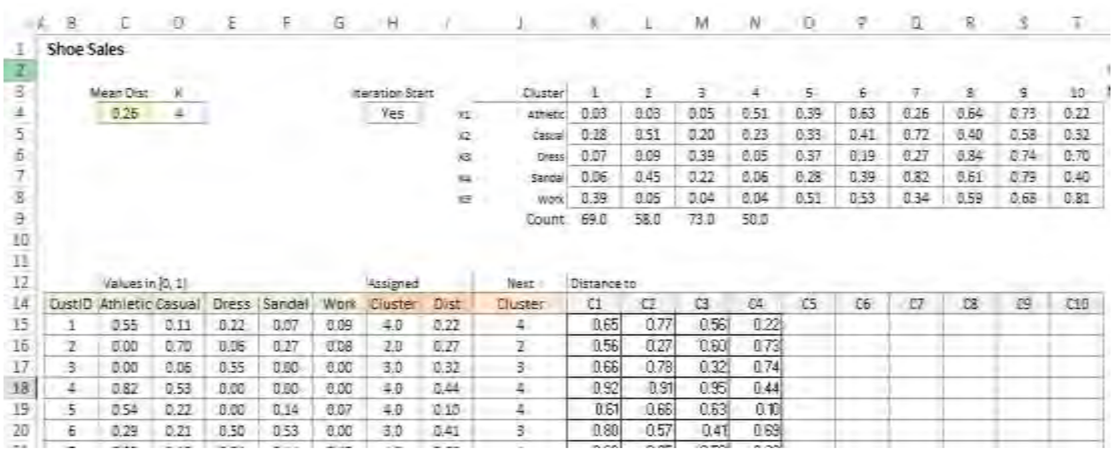


We also need to change how we assign the cluster in Column H using this iterative. If the iteration is Yes, the Cluster assigned should be the next cluster; otherwise, we use the original MOD formula in Cell, H15.

= IF($H$4="No","",J15)

We also need to compute the mean distance in C4 as =AVERAGE(I15:I264).

Figure 7: Complete solution for the K-means cluster with iterative computation for K=4



The mean distance has reduced from 0.41 to 0.26, and the Assigned cluster and the next cluster are the same after iterative computation.

Cluster one has 69 customers, and cluster two has 58, cluster 3 has 73 and cluster 4 has 50 customers.

It is essential to know the profile of the customer in each cluster. Therefore, we should also give the clusters a meaningful name based on these characteristics.

For cluster one, average spending on work and casual is higher, so we called it a casual worker cluster. The average spent on Casual is 0.28 and 0.39 on Work. For cluster two, the average spending is high on Casual and Sandal and we called it Casual sandal lover. The average spent on Casual is 0.51, it is highest among all customers and spent the most on Sandal as well with average spent of 0.45.

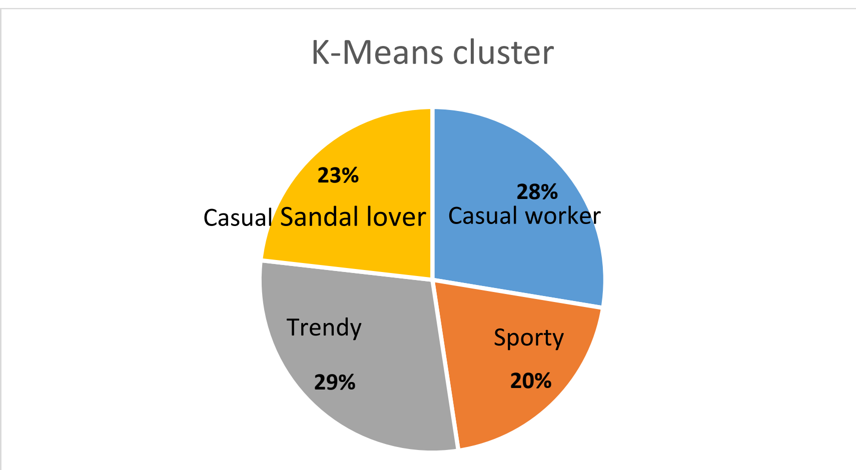
For cluster three, the average spending is high on Dress, Sandal and Casual, and we called it Trendy. For cluster four, the average spending on Athletic is highest compared to other clusters, thus we called it Sporty. The average spent on Athletics is the highest among all clusters and it is 0.51.

Figure 8 and Table 4 show the size of each cluster and % contribution for the total population.

Table 4: Cluster name and size

Cluster name	**# of customer**	**% of customer**
Casual worker	69	28%
Sporty	50	20%
Trendy	73	29%
Casual Sandal lover	58	23%

Figure 8: Pie-chart fork-mean clusters, k=4



We have run k-mean clusters using IBM SPSS modeler V18.0 by setting k=4, and we get similar results as our spreadsheets model. Figure 9 shows the size of each cluster given by SPSS, and the mean input of each cluster. E.g. for the first column, the cluster size is 29%, with 73 members in the cluster. The customer in this cluster spent 0.39 on Dress, 0.22 on Sandal and 0.20 on Casual. This will be the same as our output as Trendy.

Figure 9: The K-means clusters from IBM SPSS Modeller

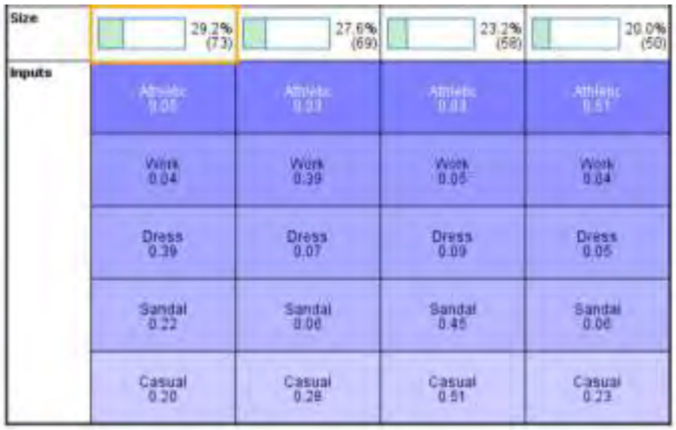
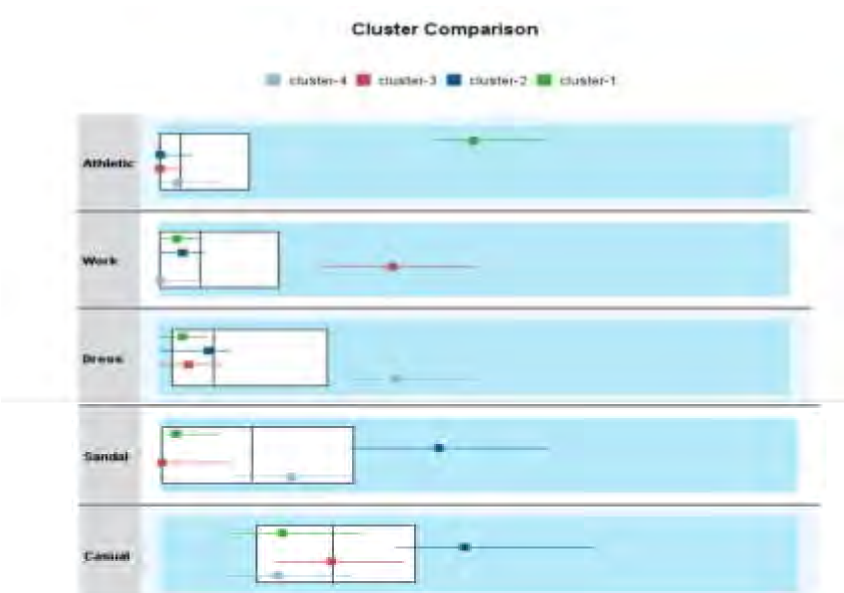


Figure 10: Output of cluster comparison from IBM SPSS Modeller



Complete model by varying k

We want to change the value of k from 2 to 10. Let us examine how to change our basic model to accommodate it.

We need to check if the cluster number is less than or equal to K given when we compute the centroid. We use the formula if cluster-id is less than or equal to K; otherwise, we will assign the considerable huge number 999 to the centroid. We change the formula in K4:K8 to the following, and we can fill up the rest of the table by dragging the formula to the right.

=IF(K$3>$D$4,999,IFERROR(AVERAGEIFS($C$15:$C$264,$H$15:$H$264,K$3), AD3))
=IF(K$3>$D$4,999, IFERROR(AVERAGEIFS($D$15:$D$264,$H$15:$H$264,K$3), AD4))
=IF(K$3>$D$4,999, IFERROR(AVERAGEIFS($E$15:$E$264,$H$15:$H$264,K$3), AD5))
=IF(K$3>$D$4,999, IFERROR(AVERAGEIFS($F$15:$F$264,$H$15:$H$264,K$3), AD6))
=IF(K$3<=$D$4,IFERROR(AVERAGEIFS($G$15:$G$264,$H$15:$H$264,K$3),AD7),999)

Distance to cluster in Cell I15, should be changed from =MIN(K15:T15)
Next cluster in Cell J15, should be changed to =MATCH(I15, K15:T15,0)

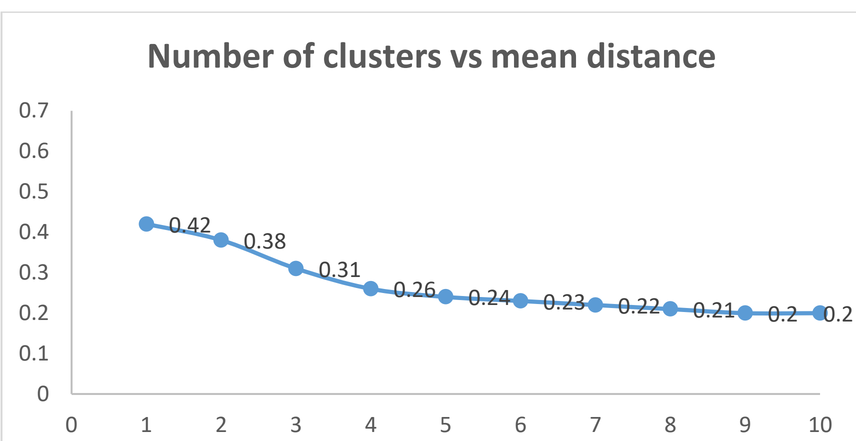
We can also drag the formula from N15 to fill the rest of the row from 015 to T15. Similarly, fill the rest of the table from 015:T264. Then we can run the entire model of varying k=1.. 10, for each k, we need to set the initial value first, set Iteration Start, H4 to No first, then set it to Yes to get the cluster assignment. This will reset the initial cluster assignment to the new number of clusters, k value, and we can get the new cluster assignment.

We look at the average distance for each k value in Table 5 and **Error! Reference source not found.**Figure 11.

Table 5: Mean distance for each k value

K, number of clusters	Mean distance
1	0.42
2	0.38
3	0.30
4	0.26
5	0.24
6	0.23
7	0.22
8	0.21
9	0.20
10	0.20

Figure 11: Number of clusters versus mean distance



From Figure 11, we can see that the mean distance reduces as we increase the number of clusters. The most significant reduction in mean distance is seen when k=2 to k=3, the mean distance has reduced from 0.38 to 0.31, and it has the steepest change, and the rate of change is 18%. From k=3 to 4, the mean distance has reduced from 0.31 to 0.26. Thus the rate of change is 16%. After k=6, the rate of change in mean distance is less than 5% and gradually becomes 0 when k=10. Thus, the best k value is 3 or 4 in this case.

Data Table method to find the best k

Instead of recording the distance for each k manually, we want to use the DATA TABLE in Excel, which will replicate the computation by changing one input cell; in this case, it will be the number of clusters, k.

We also need to ensure that we use the stratified method for the initial cluster coordinate using the Span, $AD$18 is set to 0.25. For each cluster, we multiple the span by a factor of (cluster ID)/k. If the cluster-ID is 1 and k=4, then we will have the initial cluster 1 value for all the x is, (1/4)*0.25 = 0.0625. The formula to compute this number in Excel is: (K$3/$D$4)*$AG$18

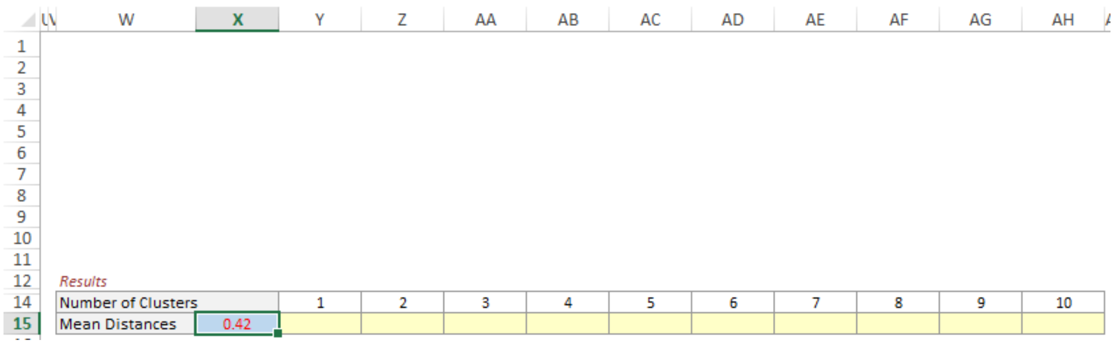
The formula to compute the coordinate in Range K4:K8 has been changed to the following:
=IF(K$3>$D$4,999,IFERROR(AVERAGEIF($H$15:$H$264,K$3,$C$15:$C$264), K$3/$D$4*$AG$18))
=IF(K$3>$D$4,999,IFERROR(AVERAGEIF($H$15:$H$264,K$3,$D$15:$D$264),K$3/$D$4*$AG$18))
=IF(K$3>$D$4,999,IFERROR(AVERAGEIF($H$15:$H$264,K$3,$E$15:$E$264),K$3/$D$4*$AG$18))
=IF(K$3>$D$4,999,IFERROR(AVERAGEIF($H$15:$H$264,K$3,$F$15:$F$264),K$3/$D$4*$AG$18))
=IF(K$3>$D$4,999,IFERROR(AVERAGEIF($H$15:$H$264,K$3,$G$15:$G$264),K$3/$D$4*$AG$18))

Figure 12: Stratified initial coordinate.

Thus, we will set up a 1D data table, where the input cell is k value from 1 to 10, and the formula is the mean distance.

Formula in Cell X15, = C4

Figure 13: Initial data table to record the mean distance



W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
Results											
Number of Clusters		1	2	3	4	5	6	7	8	9	10
Mean Distances	0.42										

Select Range from X14 to AH15, click on data what if analysis→ Data Table.
Set Row input cell is D4. Excel will change the value of D4 from 1 to 10 and record the answer given in C4 in Table 6

Table 6: Initial mean distance without repetitive iteration, Iteration is set to NO.

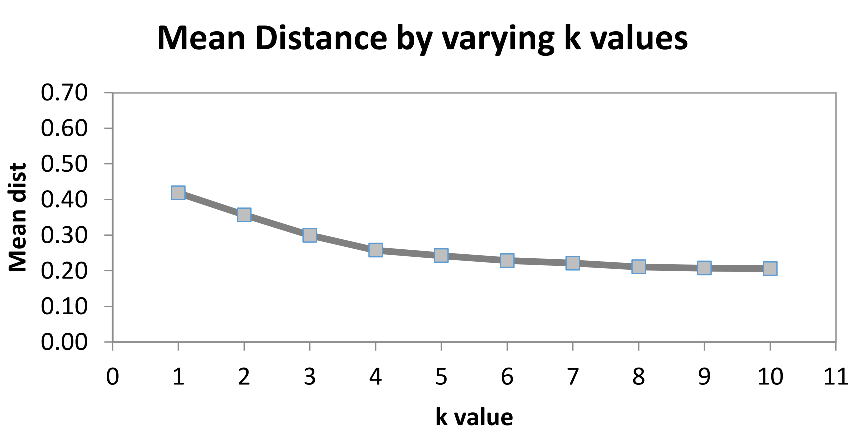
	1	2	3	4	5	6	7	8	9	10
0.42	0.47	0.43	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42

After we have set the iteration to Yes and we will get the table in Table 6, Table 7 and the graph as shown in Figure 14.

Table 7: Final mean distance result for various k values

Results											
Number of Clusters		1	2	3	4	5	6	7	8	9	10
Mean Distances	0.26	0.42	0.36	0.30	0.26	0.24	0.23	0.22	0.21	0.21	0.21

Figure 14: Mean distance versus the number of clusters k=1 to 10



CONCLUSION

In this paper, we have illustrated how to build k-mean clusters using spreadsheets without writing VBA code and using Macro. This is the first attempt in the literature where we use an innovative method to use the repetitive computations method to derive the clusters. The clustering method is a kind of unsupervised learning where we want to explore the hidden or unknown structure underlying the data and get valuable insights from the data. We first used a fixed k value, k=4, to build the basic model and then enhanced it to the advanced model by varying k values from 2 to 10 and showed that the mean distance has reduced as we increased the value of k. The same methodology can be used for other business cases, and Excel is easy to use and easily understood by users without heavy mathematics and computations.

REFERENCES

Granville V, 2017. Advanced Machine Learning with Basic Excel, Blog, Data Science Central.

Leong TY & Cheong MLF, 2015. *Business Modeling with Spreadsheets*, 3ed, 409 pages, McGraw-Hill

Leong TY & Cheong MLF, 2008. Teaching Business Modeling Using Spreadsheets, INFORMS Transactions on Education 9(1): 20–34

Leong TY & Lee WL, 2008. Spreadsheet Data Resampling for Monte-Carlo Simulation, Spreadsheet in Education 3(1):70-78

Leong TY, 2007b. Monte Carlo Spreadsheet Simulation using Resampling, INFORMS Transactions on Education 7(3):188-200

Leong TY, 2007a. Simpler Spreadsheet Simulation of Multi-server Queues, *INFORMS Transactions on Education*7(2):172-177

Ng A, 2017. Machine Learning Course, Coursera.

Roberts T, 2018. Machine Learning in Excel with Python. DataScience (https://datascienceplus.com/machine-learning-in-excel-with-python/)

Winston & Albright, 2019. Practical Management Science 6ed, Cengage.

DECISION SCIENCES INSTITUTE

Management of Platform Engineering Teams

Jordan Shropshire
University of South Alabama
jshropshire@southalabama.edu

Babalola John Omofoyewa
University of South Alabama
Bjo2021@jagmail.southalabama.edu

ABSTRACT

Platform engineering has emerged as a new function which software development organizations are beginning to manage. This function is generally aimed at reducing the cognitive burden on software developers by presenting them with a streamlined path for code publication. Platform engineers focus on integrating the various workflows and toolchains which are used to produce, test, and deploy software into production. Because this discipline is still in its infancy there is little guidance regarding its management. Hence, the purpose of this research is to formalize platform engineering by systematically defining its roles, goals, and expected outcomes. This study performs a textual analysis of articles and white papers related to platform engineering. It uses a rigorous method to conceptualize the management of platform engineering teams. The results provide a well-grounded starting point for related research. They also can also be used for assessing the performance of platform teams.

KEYWORDS: Platform engineering, DevOps, Software development, Goal alignment, Conceptual Definition

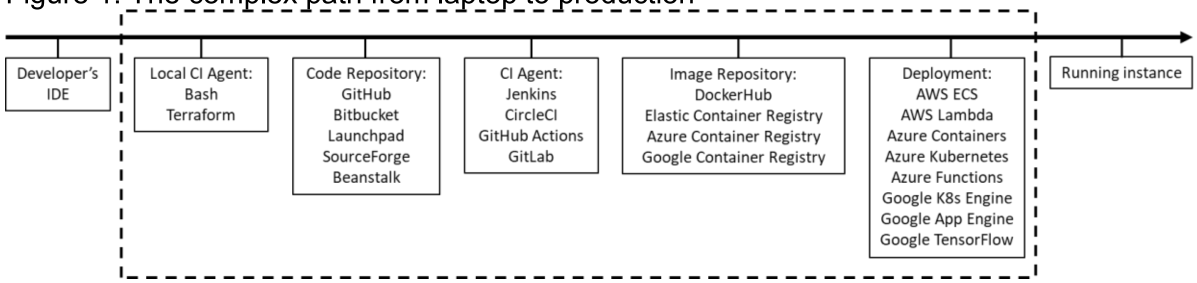
INTRODUCTION

A common misconception about software engineers is that they spend nearly all of their time producing code. However, this is not the case. A surprising amount of their time and effort is spent packaging their software into an amenable format, deploying it on company infrastructure, and supporting the associated service. Software engineering teams are commonly expected to be responsible for the entire software lifecycle. This is the essence of DevOps: holding development teams responsible for building, deploying, and running software. It has resulted in a significant shift in responsibility, with developers responsible for more and more of their application's lifecycle (Galante, 2023; Cockcroft, 2023). Software engineers are now as apt to be reconciling service issues as they are evaluating code.

In order to achieve higher levels of scalability, greater automation, and more efficient use of infrastructure, microservice architectures and technologies have become more complex (Lohrenz, 2022; Zold, 2022). Although software engineers are not generally trained to use these technologies they are expected to integrate them into their workflows. Software development teams are now expected to master a variety of additional tools such as git, continuous integration, serverless computing, and infrastructure as a service. Although these tools and services offer advantages over traditional methods they are far more complex and harder to master (Bartlow, 2022).

The role of the software engineer is far more difficult than it was just a few years ago (Schmitt, 2022). First, software engineering teams have a greater breadth of responsibility. Because of DevOps they handle everything from software development to service delivery. Furthermore, their collective operating environment is more complex (Betz, 2022). They are expected use toolsets which are more complicated and harder to troubleshoot than even just a few years ago. Taken together, these factors have significantly increased the cognitive load on engineers (see Figure1, below).

Figure 1: The complex path from laptop to production



When the cognitive burden of iterating and deploying software exceeds the engineering team's collective ability a decline in productivity is expected (Chervek, 2023). Anti-patterns are often observed. For instance, senior developers spend little time planning and developing software and more time helping junior developers process their changes. Teams may privately reallocate productive developers away from coding so they can support the operation of assigned microservices. This anti-pattern is called shadow ops, a coping model with numerous negative short-term and long-term implications. It has been suggested that teams adopting this strategy may experience a double-digit drop in overall productivity. This maladaptation is the result of an overwhelming cognitive burden placed on the team (Galante, 2022).

Forward-looking organizations have noted this trend and are taking steps to address it (Hokansan, 2020). They have stood-up teams of specialists with backgrounds in software and DevOps engineering. These groups are called platform engineering teams. The purpose of platform engineering teams is to analyze the software development and publication workflows from a holistic perspective. These teams reduce the cognitive burden by streamlining workflows and presenting “golden paths” to software developers. A golden path uses a well-integrated set of tools. It abstracts many details and presents a simplified view of the toolchain. This integrated set of tools is called the internal development platform. If executed properly it allows the software team to focus on interpreting the results of software changes instead of the mechanics of moving code into production. To sum, the purpose of the internal development platform is to reduce the cognitive load on software engineers and empower them to be more effective (Galloway, 2022). It is part of the effort aimed at reducing the cognitive burden on software teams.

The platform engineering function is still in its early stages (Garcia and Ford, 2021). However, the diffusion of this concept among software development organizations is occurring quite rapidly. This can be observed via the growth of the number of companies seeking to hire platform engineers. As more organizations launch their own teams the need for a clear articulation of platform engineering goals, roles, and expected outcomes becomes more apparent. Not only will it be necessary to standardize the role of the platform engineer but it will also be necessary to align the platform engineering team's activities with broader organizational goals (Nieman, 2020).

Hence, this research seeks to formalize platform engineering. Besides introducing the concept of platform engineering this research systematically defines its characteristic goals, its responsibilities, and expected contributions. It uses previously-validated research methods (Stemler, 2001) to conceptualize the management of platform engineering teams. Specifically, it performs a textual analysis of articles and white papers related to platform engineering. From each article it extracts reference units subjectively related to platform engineering. The units are sorted into clusters based on content similarity. The clusters are organized into more global categories according to their interrelationships. When finished, the top-level categories form an exhaustive, multi-dimensional definition of platform engineering. The resulting artifact informs the management of platform engineering teams and provides a reasonable starting point for future research.

The remainder of this research is organized as follows: The next section is the literature review. It adds context by providing background information on DevOps and the changing responsibility among software developers. The next section is the theoretical development section. It proposes an *a priori* definition of platform engineering. The following section is the methodology. It describes the process by which the conceptual definition is systematically derived. Next, the results are described. Finally, implications and concluding comments are proffered.

LITERATURE REVIEW

The purpose of this section is to provide background information in two areas. First, it explores the concept of DevOps. Second, it looks at the changing role of the software engineer.

Introducing DevOps

DevOps entered the tech world's consciousness in 2006 when the CTO of Amazon famously stated "You build it, you run it" (Galante, 2022). Up until this point software engineers were largely isolated from the day-to-day operations of their software. They had relatively little incentive to design software which was easy to deploy and maintain because they were not responsible for supporting operations. Further, they were not included in the customer feedback loop. Customer feedback that was received would have been second-hand, after being first filtered through a project or product manager. The lack of direct customer contact seemed to minimize the impetus to improve the quality of the service. The developers did not have to deal with unhappy customers.

Hence, it made sense to bridge the divide between software development and operations (Freeman-Benson and Li 2022). It was argued that the teams who create the code and are familiar with it should be the same people who fix it when an incident occurs. This would not only improve the reliability of the software and reconcile real-time issues, but it would also bring development teams closer to their customer base and hold them accountable for meeting customer requirements. DevOps reduced pressure on operations teams, forced developers to create production-ready design, and shortened response times. It also indirectly incentivized code testing. Those who have must respond to code failures at all hours are more apt to ensure that their software is well-tested and free of bugs. Collectively, these benefits provided the incentive for organizations to integrate many facets of development and operations (Gorelik, 2019).

Although organizations may vary according to their implementation of DevOps, there seems to be some consensus that the ethos is build, deploy, and support (Lohrenz, 2022, Dadgar, 2023). A single team of software engineers creates the code, performs the test, publishes the code into production, and ensures the availability of the associated service. This represents a shift away from centralized development and IT operations. The blending of skills requires teams which are well-formed and managed. There must be careful selection of team members to ensure that the appropriate mix of skill sets is available. A team which is strong in software development but weak in operational experience will tend focus more on customer feedback and less on reliability. Conversely, a team which is biased toward operational experience will focus more on reliability and less on iterating software. Well-balanced teams should be capable of designing, deploying, supporting, and updating their product from within their organization's computing ecosystem. This concept could be summarized to state that software engineering teams are responsible for delivering services, not just software (Bryant, 2022).

Changing Role of the Developer

Since its inception the concept of DevOps has remained centered on holding developers responsible for their applications' lifecycle and delivery. However, since the 2000s the development and deployment ecosystems have been increasingly complex (Garner, 2023). The workflow has become widely distributed across many disparate platforms. Developers must now shepherd their code along a convoluted path. This path begins within the integrated development environment on their computer and passes through services such as code repositories, continuous integration agents, image repositories, image registries, and infrastructure-as-a-service platforms prior to deployment. Software teams may be required to create Git repositories, Docker images, Jenkins templates, Terraform code, Cloud CLIs, and bash scripts in addition to developing software. These tools can be very useful. Many of them provide a high level of automation. However, it can be difficult to predict the behavior of automated tools. It is even harder to identify and isolate problems when multiple automation tools interact with each other (Horovitz, 2022). To sum, the DevOps work environment has become increasingly complex.

The increased complexity of the current DevOps environment has added to the cognitive load on software engineers (Dodd, 2023). Cognitive load is the amount of cognitive processing capacity required to complete a task (Sweller, 1988). A task with a higher cognitive load requires more cognitive processing. However, people have finite processing capacities. Thus, when a task requires more cognitive processing it means that less capacity remains for other problem-solving aspects such as schema acquisition (Goncales, Farias, da Silva, 2021). Individuals can become overwhelmed by large cognitive loads and struggle to complete ancillary tasks which they could otherwise execute without trouble. A well-documented effect of increased cognitive loads is decreased learning ability (Abbad-Andaloussi, 2023). The individual overwhelmed by their new workspace will have difficulty learning, understanding, and mastering it. They will struggle to keep up with their work. This has been overserved in many fields, including within the DevOps computing field (Tsang, 2023). As the DevOps technical environment has grown in complexity so has the cognitive load. And the increase in cognitive load leads to decreases in the performance of software engineers.

Generally speaking, engineering managers have multiple options for helping their teams cope with heavy cognitive burdens (Zold, 2021). One approach is to silo organizational functions and allow for specialization. In this role an individual has less breadth of responsibility. Instead they become expert in processing work and solving problems within a very narrow band of topicality. In this scenario the cognitive processes required to complete tasks align sufficiently to allow the worker to gain deeper insights into their domain. Another coping strategy is to maintain the breadth of responsibility but simplify the expected output. The individual must still be familiar with a breadth of tools. However, the output is less troubling. A less complex product requires less attention to detail, thus alleviating the cognitive load. Another method is to simplify the tool set. If the manager can take away tools, cut out processing steps, or otherwise simplify the workflow. This shrinks the cognitive burden associated with mastering a plurality of tools and allows the worker to focus on improving their product (Smiler, 2022). Managers can use any combination of these approaches to reduce the cognitive load on their teams.

Of the aforementioned approaches, platform engineering is most akin to the latter approach (Delp, Gracely, and Hirschfeld, 2023). It reduces cognitive load by presenting a more seamless workflow so that software engineers are free to focus on their end product. Platform teams focus on building golden paths, streamlined avenues to code development and deployment. They make use of abstraction, integration, and automation. Their primary mission is to make it easier for software teams to get their work done. To summarize, the platform engineering function allow software teams to concentrate on the quality of their services because it reduces the cognitive load associated with the DevOps toolset (Gottardo, 2022).

TOWARDS A CONCEPTUAL DEFINITION

Savvy managers will strive ensure that their platform teams meet their objectives without simply moving the cognitive burden onto their own shoulders (Thiruvengadam, 2023). They will consider the needs of software teams and the capacity of platform teams. This requires an understanding of the general expectations and responsibilities of platform teams. While some of the more technical aspects of platform engineering have been researched there is relatively little guidance in the management and information systems literature to assist those responsible for leading platform teams. The literature which presently exists is often incomplete on its own and inconsistent with other parts of the literature.

One area which is often confused in the literature is the difference between DevOps and platform engineering. The purpose of DevOps is to manage the tools and infrastructure for software deployment. Its primary emphasis is on automation: ensuring continuous integration of software into production (Hokanson, 2020). The purpose of platform engineering is present a streamlined toolchain to software developers. Its emphasis is on integration: combining a number of discrete tools into a single, harmonious development platform. Whereas DevOps focuses on tools platform focuses on people. These functions often overlap. For instance, a platform engineer cannot align DevOps tools if those tools are not available.

A second area of confusion is the relationship between site reliability engineering and DevOps engineering. Site reliability engineers use DevOps tools to automate IT infrastructure tasks such as system health management and application monitoring (Dean, 2022). They commonly address operational, scalability, and reliability issues. Site reliability engineering seeks to improves the stability and quality of services as they grow in scale. By contrast, DevOps engineers solve problems related to the software deployment pipeline (Sandilands, 2022).

Whereas site reliability is typically more development oriented the DevOps discipline is more ops and IT focused.

Although these functions may seem redundant they bring different perspectives which ensure that reliable, quality services are rendered (Villanueva, 2022). A single organization could have multiple software engineering teams which create code and deploy it in the form of a microservice. The organization would require DevOps engineers to establish tools and infrastructure for deployment. The organization could use site reliability engineers to ensure the scalability of services as scale. It would also use platform engineers to integrate the various tools and dashboards into a single internal development platform so that software updates are easily delivered. An appreciation of these nuances is an important first step in understanding and managing platform engineering.

Thus, this study reviews the extant literature and develops a conceptional definition of platform engineering. In doing so it addresses the goals, tasks, and expected outcomes of platform engineering teams. The results will not only lay the groundwork for further research but also inform managers seeking to establish platform teams.

METHODOLOGY

This purpose of this section is to systematically define platform engineering and articulate the facets which would be relevant to management and research. This definition will conceptualize the goals and the roles of the platform engineering function. The method used to derive this artifact is content analysis of textual inputs (Drisko and Maschi, 2016). Content analysis is a research method from the social sciences which draws inferences from text (Weber, 1990). In this study, the text will include articles related to platform engineering. Each reference to an aspect of platform engineering is classified according to an a priori coding scheme. The investigative team members independently classified each reference unit and later convened to compare results. The initial level of agreement was 81% of cases. In instances in which there was disagreement the researchers collectively reviewed the attributes of reference units until a consensus was achieved. In two cases an independent researcher with experience in cloud computing broke a tie. The results of the coding operation were iteratively refined. First, reference units were clustered into groups based on apparent similarity. Next, these clusters were organized into categories. Each category represents a separate facet of the conceptual definition of platform engineering.

The resulting artifact is expected to be of value to both managers and researchers. It provides value at two levels of analysis. At a global level, the definitional components explain the high-level attributes of platform engineering, such as its purpose, goals, and expected outcomes. At a granular level, the definitional components outline specific tasks and responsibilities. This method of qualitative research is regularly used within the social sciences and computing fields (Templeton et al., 2002). For instance, it is often used by information systems researchers to define concepts and frameworks for which little research currently exists (e.g., Byrd and Turner, 2000; Lewis et al. 2005). Hence, it is expected that the outcome will be utility in future research and in practice.

Sample

This research used text from articles, whitepapers, and professional blog postings about platform engineering. The sample was drawn from this population as follows: academic publication databases, academic library catalogs, and search engines were queried using the keywords “platform engineering.” Over 139 unique writings were identified. After an initial inspection it was determined that 77 articles did not contain content which was related to the purpose of the study. The majority of these articles used the phrase “platform engineering” to reference to the engineering of an actual mechanical platform or they used the term ‘platform’ as a generic reference to a vehicle or other systems. These articles were associated with journal of mechanical, systems, and aerospace engineering. An additional 14 articles did not contain sufficient details for extracting reference units. This left a total of 48 articles from which aspects of platform engineering could be gleaned (see Appendix A). It should be noted that few of the articles were from academic sources. The majority were from whitepapers, magazines, online articles, and other practitioner-related sources. Most were written for audiences in the software development field.

Selection of Reference Units

Each article was reviewed for specific references to the platform engineering function. Each individual reference is classified as a reference unit. Each reference unit is a specific reference to one aspect of platform engineering. All references were treated as unique data to code. An a priori coding scheme was used to code the data (Stemler, 2001). The coding scheme was initially based on a 24-item list of responsibilities contained in a recruiting firm’s advertisement for a platform engineering manager (UCAS, 2020). This list was unique in that it did not prescribe specific treatments or actions but instead described its various activities. This was used to classify the initial 10% of reference items. After which, the authors compared their results and expanded the coding scheme in order to clarify items which captured multiple concepts in a single item. When complete, the list grew to a scheme of 31 codes. Although this method is subject to the same biases as other qualitative techniques it is still with frequency in content and textual analysis studies (Hardwood and Garry, 2003). The updated coding scheme was reapplied to the reference units. At the 50% point the updated coding scheme was re-evaluated and confirmed.

Category Development

All told, 370 reference units were extracted from the articles included in the final sample. The referenced units were coded. The data was coded into a series of 33 of platform engineering management activities. These 33 activities were clustered into 12 independent groups based on apparent similarity. As with coding, clustering is a qualitative technique which introduces some degree of subjectivity. Hence, a standardized process was used to ensure an appropriate degree of rigor (Weber, 1990). The technique by which the categories were created consisted of three phases (Krippendorf, 1980). First, the activities which were most similar were identified. Combining these activities would have the smallest effect on the observed differences in the data as a whole if they merged. Thus in the third step the activities were merged, creating new-found clusters. This three-step process repeated until no further mergers could be conducted. Once the clusters were completed they were given a name which appropriately described their common traits. These named clusters formed the basis of the final 12 categories. Each of the categories was then labeled according to its contribution to management. Each category was then also labeled as either a goal, a task, or expected outcome of platform engineering.

RESULTS

As a result of the textual analysis, clustering, and category definition, a conceptual definition of platform engineering management was created (see Table1). The role of platform engineering is to create automated tools and workflows. The expected outcome is the delivery of a development platform which meets user requirements. The managerial tasks include creating a platform engineering roadmap, creating a technology roadmap, working with others to reduce or eliminate tasks, anticipate project features, ensure infrastructure suitability, provide platform expertise, develop and document a centralized API estate, define and manage platform policies, provide in-depth evaluation, and engage the platform team. Although the articles in the textual analysis suggested specific management actions, this definition is more generally defined in order to assure its global applicability.

Table 1: Conceptual Definition of Platform Engineering

	Category	Activities
Goals	Create tools and automated workflows:	Allow software engineers to test & deploy at maximum velocity
		Enable a golden path for code publication
Expected Outcomes	Deliver a platform which meets requirements:	Ensure the platform is flexible and consistent
		Ensure the platform delivers the tools, interfaces, and components to drive products & services
Managerial Tasks	Create a platform engineering roadmap:	Ensure that platforms & products remain current
		Ensure that platforms & products deliver value to product teams
		Ensure that platforms & products support stakeholders in terms of scalability, resilience, capacity, and operational maintenance
	Create a technology roadmap:	Ensure continuous improvement of supported solutions
		Ensure technologies empower a DevOps culture
		Ensure integration of automated platform capabilities
	Work with others to reduce or eliminate tasks:	Collaborate with product engineering
		Collaborate with service management
		Collaborate with operations
	Anticipate product features:	Track dependent product strategies and roadmaps
		Elicit stakeholder input
		Monitor industry advances
		Collect business needs
	Ensure infrastructure suitability:	Ensure infrastructure supports multiple platforms
		Confirm that infrastructure maintains high levels of resilience, performance, and capacity
	Provide platform expertise:	Provide subject expertise to other engineering teams
		Consult to management
		Mentor members of the platform engineering team
	Develop and document a centralized API estate:	Promote API reuse
		Support API versions
		Ensure end-to-end lifecycle management
	Define and manage policies:	Manage the platforms' security
		Control the flow of data across platforms
	Perform continuous evaluation:	Scan for evolving technologies
		Identify mature technologies
		Track decline of legacy technologies
	Engage the platform team:	Task platform members to execute the platform engineering roadmap
		Provide measures to assess team quality and effectiveness
		Ensure clarity of purpose and deliverables

IMPLICATIONS AND CONCLUSION

Platform engineering's lineage can be traced back to the DevOps movement. It was spurned by the expectation that product teams will deploy and support the software they create. Its emergence as a new discipline is a response to the increased cognitive workloads on software engineers. As the ecosystem grows and the tools became more complex it becomes harder to sustain software development. To counter these complexities and improve the effectiveness of product teams some organizations have engaged teams of engineers to simplify the software development and publication processes.

This research undertook a structured examination of the platform engineering function and contrived a conceptual definition of its goals, expected outcomes, and tasks. It uses qualitative research methods to identify textual samples, extract and classify related data units, and merge them into higher-order categories. It finds that the goal of software engineering is to provide a simplified workflow – a golden path – which begins at the developer's laptop and ends at the point of software execution within a public cloud environment. It is expected that platform engineering teams will integrate the various toolsets and yield an internal development platform. Managerial tasks associated with platform engineering include creating platform roadmaps, continually assessing technologies, coordinating with other engineering teams and defining policies. Platform engineering managers can use the proposed artifact as a checklist to ensure they are fully executing their duties.

Platform engineering is unique in several respects. First, if it one of the few business functions which focus on enabling others to do their job more effectively. It is not common for software development organizations to allocate resources to units which are not directly related to product development or deployment. Second, it recognizes that the cognitive workload among software engineers is approaching the point of saturation. It is expected that high levels of employee burnout and turnover coincide with the decision to launch platform teams. Third, although platform engineering requires a technical background most of the work involves coordinating and communication. Management of this engineering function requires soft skills and an appreciation of psychology and human factors. It is expected that in the future research will focus on developing tools and methods to measure and reduce cognitive load.

APPENDIX

Table 1: Data Sources Used in Content Analysis	
Bartlow, 2022	Horovits, 2022
Betz, 2022	Johnson, 2023
Bridgwater, 2023	Lohrenz, 2022
Bryant, 2022	Majors, 2022
Chervek, 2023	Markova, 2023
Cockcroft, 2023	Mink, 2022
Coffman, 2020	Morrissey, 2023
Dadgar, 2023	Naumovska, 2021
Dagher, 2022	Nieman
Dean, 2022	Perri, 2022
Delp, Gracely, Hirschfeld, 2023	Perry, 2022
Ditiangkin, 2022	Sandilands, 2023
Dodd, 2023	Sandilands, 2023
Doerrfeld, 2022	Schmitt, 2022
Doerrfeld, 2023	Schults,2021
Freeman-Benson, Li, 2022	Smiler, 2022
Galante, 2022	Sonpatki, 2022
Galante, 2023	Thiruvengadam, 2023
Galloway, 2022	Tsang, 2023
Garcia, Ford	Tunggal, 2022
Garner, 2023	Villanueva, 2022
Gorelik, 2019	Walker, 2023
Gottardo, 2022	Zold, 2023
Hokansan, 2020	Zold, 2022

REFERENCES

Bartlow, A. (2022) Platform Engineering vs. Site Reliability Engineering. Retrieved from: https://www.aha.io/engineering/articles/platform-engineer-vs-site-reliability-engineering February 21.

Betz, C. (2022) Platform Product Management Versus Platform Engineering. Retrieved from: https://www.forrester.com/blogs/platform-product-management-versus-platform-engineering/ March 11

Bridgwater, A. (2023) Retrieved from: https://www.forbes.com/sites/adrianbridgwater/2023/01/20/levelling-up-time-for-it-platform-teams/?sh=7182fb25313a February 21.

Bryant, D. (2022) Levelling Up Time for IT Platform Teams Retrieved from: https://www.getambassador.io/kubernetes-expert-interviews/self-service March 11

Byrd, T., Turner, D. (2000) Measuring the flexibility of information technology infrastructure: exploratory analysis of a construct. Journal of Management Information Systems, 17(1) 167-208.

Chervek, E. (2023) How Platform Engineers Help Narrow Cloud Skills Gap. Retrieved from: https://www.sdxcentral.com/articles/interview/how-platform-engineers-help-narrow-cloud-skills-gap/2023/02/ March 11

Cockcroft, A. (2023) Platform Engineering Teams Done Right…Retrieved from: https://adrianco.medium.com/platform-engineering-teams-done-right-b3b3d4a8ad23 March 6

Coffman, S. (2020) Advocating for a Product Mindset within Platform Teams and How We Do It at HelloTech (Part 1). Retrieved from: https://engineering.hellofresh.com/advocating-for-a-product-mindset-within-platform-teams-and-how-we-do-it-at-hellotech-part-1-fc1fbf8ae015 March 6

Dadgar, A. (2023) What Is a Platform Team and What Problems Do They Solve? Retrieved from: https://www.hashicorp.com/resources/what-is-a-platform-team-and-why-do-we-need-them March 6

Dagher, K. (2022) 8 Ways to Build Platform Engineering Teams. Retrieved from: https://fellow.app/blog/leadership/ways-to-build-platform-engineering-teams/ March 6

Dean, M. (2022) The Platform Team, A Must Adopt Approach? Retrieved from: https://www.udig.com/digging-in/platform-team-must-adopt-approach/ February 21.

Delp, A., Gracely, B., Hirschfeld, R. (2023) 2023 Look Ahead to Platform Engineering. Retrieved from: https://www.thecloudcast.net/2023/01/2023-look-ahead-to-platform-engineering.html March 11

Ditiangkin, L (2022) Retrieved from: https://platformengineering.org/talks-library/a-shared-vision-for-platform-engineering March 6

Dodd, R. (2023) What is Platform Engineering? Building a Shared Vision. Retrieved from: https://launchdarkly.com/blog/what-is-platform-engineering/ March 6

Doerrfeld, B. (2022) How Tech Leaders Can Reduce Workplace Burnout. Retrieved from: https://accelerationeconomy.com/cybersecurity/how-tech-leaders-can-reduce-workplace-burnout/ March 11

Doerrfeld, B. (2023) Strategies To Consider When Adopting Platform Engineering. Retrieved from: https://devops.com/strategies-to-consider-when-adopting-platform-engineering/ March 6

Drisko, J., Maschi, T. (2016) Content Analysis, Oxford University Press, Oxford, UK
Freeman-Benson, Li, R. (2022) SRE vs. Platform Engineering Retrieved from: https://www.getambassador.io/resources/rise-of-cloud-native-engineering-organizations March 11

Galante, L. (2022) What is platform engineering? Retrieved from: https://humanitec.com/platform-engineering#how-do-you-build-a-platform-engineering-team March 11

Galante, L. (2023) What is platform engineering? Retrieved from: https://platformengineering.org/blog/what-is-platform-engineering, February 21.

Galloway, M. (2022) Your Platform Org needs a purpose. Here's how to find it. Retrieved from: https://medium.com/@michael.roy.galloway/your-platform-org-needs-a-purpose-heres-how-to-find-it-64874b082d80

Garcia, C., Ford, C. (2021) Mind the platform execution gap. Retrieved from: https://martinfowler.com/articles/platform-prerequisites.html March 6

Garner, R. (2023) What is an Engineering Management Platform? Retrieved from: https://linearb.io/blog/engineering-management-platform/, February 17

Gorelik, S. (2019) Nuts and bolts of building a platform team (Flatiron Health). Retrieved from: https://www.datadoghq.com/videos/nuts-bolts-of-building-a-platform-team/

Gottardo, A. (2022) Driving value for our clients, via platform engineering. Retrieved from: https://www.osedea.com/en/blog/osedea-platform-engineering March 6

Harwood, T., Garry, T. (2003) An overview of content analysis. The Marketing Review 3(4), 479-498.

Hokansan, E. (2020) The Rise of Platform Engineering. Retrieved from: https://softwareengineeringdaily.com/2020/02/13/setting-the-stage-for-platform-engineering/ March 11

Horovits, D. (2022) Platform Engineering: DevOps evolution or a fancy rename? Retrieved from: https://horovits.medium.com/platform-engineering-devops-evolution-or-a-fancy-rename-a0cad2cdc819 February 21.

Johnson, B. (2023) Top 10 Platform Engineering Tools. Retrieved from: https://linearb.io/blog/platform-engineering-tools/

Krippendorff, K. (1980) Content analysis: An introduction to its methodology, Sage, London.
Lewis B., Templeton, G., Byrd, T. (2004) A methodology for construct development in MIS research. European Journal of Information Systems, 14, 2, 388-400

Lohrenz, S. (2022) Platform Engineering: What Is It and Who Does It? Retrieved from: https://thenewstack.io/platform-engineering/platform-engineering-what-is-it-and-who-does-it/ March 6

Majors, C. (2022) The Future of Ops Is Platform Engineering. Retrieved from: https://www.honeycomb.io/blog/future-ops-platform-engineering March 6

Markova, S. (2023) 7 Platform Engineering KPIs you should be tracking. Retrieved from: https://blog.cycloid.io/7-platform-engineering-kpis-you-should-be-tracking February 21.
Mink, J. (2022) On Building a Platform Team. Retrieved from: https://www.honeycomb.io/blog/building-platform-team March 6

Morrissey, D. (2023) The Future Is Platform Engineering. Retrieved from: https://www.f5.com/company/blog/the-future-is-platform-engineering March 6

Naumovska, M. (2021) What Does A Platform Team Do? Explaining the Challenges They Face. Retrieved from: https://medium.com/geekculture/what-does-a-platform-team-do-explaining-the-challenges-they-face-ef6c0d904e36 March 6

Nieman, G. (2020) How We Use Golden Paths to Solve Fragmentation in Our Software Ecosystem Retrieved from: https://engineering.atspotify.com/2020/08/how-we-use-golden-paths-to-solve-fragmentation-in-our-software-ecosystem/ March 6

Perri, L. (2022) What's New in the 2022 Gartner Hype Cycle for Emerging Technologies. Retrieved from: https://www.gartner.com/en/articles/what-s-new-in-the-2022-gartner-hype-cycle-for-emerging-technologies February 24

Perry, M. (2022) Platform Engineering: Explained, Benefits, and How to Get Started? Retrieved from: https://www.qovery.com/blog/organizationsplatform-engineering-explained-benefits-and-how-to-get-started-are-embracing-platform-engineering-because February 24

Sandilands, D. (2023) Platform Engineering vs. DevOps: Is There a Difference? Retrieved from: https://www.puppet.com/blog/platform-engineering-vs-devops February 24

Sandilands, D. (2023) What is Platform Engineering? Retrieved from: https://www.puppet.com/blog/what-platform-engineering

Schmitt, J. (2022) What is platform engineering? A quick introduction. Retrieved from: https://circleci.com/blog/what-is-platform-engineering/ February 24

Schults, C. (2021) What Is Platform Engineering? The Concept Behind the Term. Retrieved from: https://www.liatrio.com/blog/what-is-platform-engineering-the-concept-behind-the-term February 24

Smiler, N (2022) The Six Pillars of Platform Engineering Retrieved from: https://www.fpcomplete.com/platformengineering/ February 24

Sonpatki, P. (2022) The difference between DevOps, SRE, and Platform Engineering. Retrieved from: https://last9.io/blog/devops-vs-sre-vs-platform-engineering/ March 6

Stemler, S. (2001) An overview of content analysis. Practical Assessment, Research, and Evaluation: Vol. 7, Article 17.

Templeton, G. Lewis, B., Snyder, C. (2002) Development of a measure for the organizational learning construct, Journal of Management Information Systems, 19, 2, 175-218.

Thiruvengadam, V. (2023) How to Get Platform Engineering Team Structure and Skills Right. Retrieved from: https://techspective.net/2023/02/21/how-to-get-platform-engineering-team-structure-and-skills-right/ March 6

Tsang, D. (2023) How Platform Engineering Can Help Accelerate Release Velocity and Quality. Retrieved from: https://saucelabs.com/resources/blog/how-platform-engineering-can-help-accelerate-release-velocity-quality March 6

Tunggal, A. (2022) Platform Engineer Job Description. Retrieved from: https://himalayas.app/advice/platform-engineer-job-description March 6

Villanueva, S. (2022) What is platform engineering? Retrieved from: https://napptive.com/blog/what-is-platform-engineering/ March 6

Walker, J. (2023) What Is Platform Engineering? Role, Principles & Benefits. Retrieved from: https://spacelift.io/blog/what-is-platform-engineering March 6

Weber, R. (1985) Basic content analysis, Sage, London.

Zold, G. (2021) Managing Platform Teams: How to Build and Structure Platform Engineering? Retrieved from: https://codingsans.com/blog/managing-platform-teams March 6

Zold, G. (2022) Platform Engineering 101: Takeaways from Building Internal Tools at HubSpot. Retrieved from: https://codingsans.com/blog/platform-engineering March 6

DECISION SCIENCES INSTITUTE

Mediating Role of Internal Integration in Business Analytics Capability's Impact on Supply Chain Performance

Canchu Lin
SUNY Farmingdale
linc@farmingdale.edu

Andrea Xu
SUNY Farmingdale
xul1@farmingdale.edu

ABSTRACT

This study investigates internal integration as an outcome of practicing business analytics capability and its mediating role in influencing a firm's supply chain performance indicated by relationship building, knowledge co-creation, and new product development. It proposes a research model with propositions regarding relationships among the concepts.

KEYWORDS: Business analytics capability, Internal integration, Relationship building, knowledge co-creation, and new product development

DECISION SCIENCES INSTITUTE

Modelling ERP Back Order Processing in Spreadsheets

Anil Singh
Department of Information Systems & Operations Management
University of Texas at Arlington
E-mail: anil.singh@uta.edu

Vikram Bhadauria
Texas A&M University-Texarkana
Office: SCIT 114
Email: vbhadauria@tamut.edu

Lyseth D'Souza
National Institute of Construction Management and Research, Shamirpet, Hyderabad, Telangana, India
Email: ldsouza@nicmar.ac.in

Vimlesh Prabhu Desai
National Institute of Construction Management and Research, Shamirpet, Hyderabad, Telangana, India
Email: vprabhudesai@nicmar.ac.in

ABSTRACT

Backorder Processing is the distribution of goods that are in shortage due to high demand or low supply. Depending on certain conditions, customers are categorized into in-built categories (strategies) where important customers may gain and less important may stand to lose. To extend student knowledge in this area we develop a teaching method where students complete the distribution manually and then automate the process by developing a spreadsheet-based model. Doing the process manually and replicating the process in spreadsheets –makes it easier for students to understand. automated backorder processing in ERP systems.

KEYWORDS: Spreadsheet, Backorder Processing, ERP

INTRODUCTION

ERP education in Business

There are benefits to incorporating ERP in business education. To this end business schools have covered various ERP related concepts in business curriculum (Hepner & Dickson, 2013; Vician, & Mortenson, 2017). ERP education in business education is crucial as students can understand better business concepts such as the HR (Hawking, Foster, & Bassett, 2002), accounting cycle and business transaction processing (Bae & Lee, 2021), internal controls (Gainor, Bline, & Zheng, 2014), segregation of duties (Barkhi, & Kozlowski, 2017; Jones & Mensching, 2007) and audit (Debreceny et al 2005).

Business Process Automation in ERPs

A big reason for ERPs to become so powerful is automation. For example, the automated pricing in the ERP system calculates values of an order by extracting prices based on quantity, discounts, and taxes from the table to populate the sales orders and obtain the sales order total. (Singh, Bhadauria, & Mangalaraj, 2023). Similarly, automation happens in invoicing (Danner et al. 2021) and in payments and in dunning.

Back Order Processing

One area of the areas of automation is backorder processing. *" The basic idea of backorder processing is to carry out a new ATP check for a set of order items. This way backorder processing can also be used to distribute quantities in case of shortage (resp. lateness) according to priorities.” (*Dickersbach & Dickersbach, 2009). Such a situation can happen due to high demand or low supply. The backorder processing can be used to “reconfirm” sales orders. Hence, in backorder processing, you may end up refusing (overriding) a confirmed order from one customer and give it to another customer. Backorder processing can be set up for a combination of delivery priority, item category and sales organization (segment). Table 1 provides a representative list of shortages in recent years.

Table1: Representative list of material shortages

MATERIAL	REFERENCE
Ventilators and personal protective equipment	Ranney, Griffeth, & Jha (2020).
Computer Chip Shortage	Casper et. al. (2021)
Toilet Paper Shortage	Paul & Chowdhury, (2020)
Baby formula Shortage	Doherty et. al. (2022)
Integrated Circuits	Pennisi (2022)
Personal Protective Equipment	Burki (2020)
Helium Shortage	Reisch (2012)

TEACHING MODULE

In line with exercises that have focused on hands-on exercises (Zadeh et al., 2020) in this module, the students are first required to provide four industry examples of material shortages (1 point). Second, the students are asked why an organization would need to prioritize and rank its customers (rank) (4 reasons) (2 points). Third, In the manual method, students are provided with five scenarios (5 situations – 5 points) where shortages exist - and asked to calculate the units given to each confirmation strategy category (priority category) based on table 1 of student handout in the Appendix. Finally, the students are required to automate the process by creating a spreadsheet model that will automatically allocate the material to various confirmation strategies (priority categories). Students will use spreadsheets to model the above mentioned and submit the file to instructor (2 points).

Method Efficacy

In the pilot study of this method, students were surveyed after completion of the assignment. The following questions adapted from Davis and Comeau (2004) and Zadeh et al., (2020). Table 2 presents the results of the student feedback. IRB approval is in process.

Table 2: Student Survey Feedback (N=7)

Variables	Mean	Std. Deviation
My knowledge of backorder processing before doing this exercise was: (1: little to 5: high)	2.25	1.25
This exercise improved my understanding of distribution of items when there is a shortage (1: Strongly disagree to 5: Strongly agree)	5	0
The exercises were reasonable and useful (1: Strongly disagree to 5: Strongly agree)	2.5	0.47
The steps described in the handout were not working (1: Strongly disagree to 5: Strongly agree)	1	0
I gained no new knowledge from these exercises (1: Strongly disagree to 5: Strongly agree)	1.75	1.41
The student handout was easy to follow (1: Strongly disagree to 5: Strongly agree)	4.75	0
The debriefing added value to my understanding of automation of allocation during shortage (1: Strongly disagree to 5: Strongly agree.)	4.75	0.94
How much time did you take to complete this assignment (in minutes)	33.75	0.94
Assessment Results Average (10 points)	9.4	

Debriefing

After the assignment, the instructor can apprise students of the connection between automation and ERPs. The instructor can reveal examples of shortages and the various reasons for ranking the customers. The instructor can then show the manual answers, show the correct modelling technique in spreadsheets. The spreadsheet model is shown (Appendix). We mark the cells based on decisions to be made. Green meaning- must send material; Blue means- depends on availability and Red means – cancel/postpone orders.

The instructor can go into deeper details regarding the Back Order Processing. For example, if more than one segment (selected customers based on certain criteria) is assigned to the same confirmation strategy, the sequence of the segments within the confirmation strategy is used as a prioritizing method. Furthermore, the system can be set to raise an exception if the order is not fully confirmed.

Conclusion

The process of prioritizing sales order during shortages is called backorder processing. This teaching module shows manual method and spreadsheet method of the backorder process in ERPs. The average completion time during the pilot testing was 34 minutes. Universities with access to SAP ERP can show the process in the module. This will help students understand the technical and non-technical aspects of backorder processing.

References

Bae, B., & Lee, C. C. (2021). Using ERP to Teach Accounting Courses. International Journal of Accounting & Finance Review, 8(1), 1-5.

Barkhi, R., & Kozlowski, S. (2017). ERP in the Classroom: Three SAP Exercises Focused on Internal Controls. Journal of Emerging Technologies in Accounting, 14(1), 77-83.

Burki, T. (2020). Global shortage of personal protective equipment. *The Lancet Infectious Diseases*, *20*(7), 785-786.

Casper, H., Rexford, A., Riegel, D., Robinson, A., Martin, E., & Awwad, M. (2021, August). The impact of the computer chip supply shortage. In Proceedings of the international conference on industrial engineering and operations management, Bangalore, India (pp. 236-245).

Danner, M., Maurer, B., Schuh, S., Greff, T., & Werth, D. (2021). Invoice Automation: Increasing Efficiency in the Office at Satherm GmbH Using Artificial Intelligence. Digitalization Cases Vol. 2: Mastering Digital Transformation for Global Business, 45-60.

Davis, C. H., & Comeau, J. (2004). Enterprise integration in business education: Design and outcomes of a capstone ERP-based undergraduate e-business management course. Journal of Information Systems Education, 15(3), 287-300.

Debreceny, Roger S., Glen L. Gray, Joeson Jun-Jin Ng, Kevin Siow-Ping Lee, and Woon-Foong Yau. "Embedded audit modules in enterprise resource planning systems: Implementation and functionality." Journal of Information Systems 19, no. 2 (2005): 7-27.

Dickersbach, J. T., & Dickersbach, J. T. (2009). Order Fulfilment Overview. Supply Chain Management with SAP APO¿ Structures, Modelling Approaches and Implementation of SAP SCM¿ 2008, 97-98.

Doherty, T., Coutsoudis, A., McCoy, D., Lake, L., Pereira-Kotze, C., Goldhagen, J., & Kroon, M. (2022). Is the US infant formula shortage an avoidable crisis?. The Lancet, 400(10346), 83-84.

Gainor, M., Bline, D., & Zheng, X. (2014). Teaching internal control through active learning. Journal of Accounting Education, 32(2), 200-221.

Hawking, P., Foster, S., & Bassett, P. (2002). An applied approach to teaching HR concepts using an ERP system. In The Proceedings of Informing Science and IT Education Conference, Cork, Ireland, Informing Science 699-704.

Hepner, M., & Dickson, W. (2013). The value of ERP curriculum integration: perspectives from the research. Journal of Information Systems Education, 24(4), 309.

Jones, N., & Mensching, J. (2007). A segregation of duties case study in the purchase-to-pay process with an SAP example. AIS Educator Journal, 2(1), 1-21.

Paul, S. K., & Chowdhury, P. (2020). Strategies for managing the impacts of disruptions during COVID-19: an example of toilet paper. Global Journal of Flexible Systems Management, 21, 283-293.

Pennisi, S. (2022). Pandemic, Shortages, and Electronic Engineering. IEEE Circuits and Systems Magazine, 22(3), 41-49.

Ranney, M. L., Griffeth, V., & Jha, A. K. (2020). Critical supply shortages—the need for ventilators and personal protective equipment during the Covid-19 pandemic. New England Journal of Medicine, 382(18), e41.

Reisch, M. S. (2012). Helium shortage affecting instrument users to extend into 2013. Chemical and Engineering News, 90(29), 32-34.

Singh, A., Bhadauria, V. S., & Mangalaraj, G. (2023). Teaching Tip: A Teaching Module Illustrating ERP Item Value Automation. Journal of Information Systems Education, 34(1), 16-31.

Vician, C., & Mortenson, K. G. (2017). Integrating Accounting Themes with Enterprise System Experiences: An Accounting Department's Journey. AIS Educator Journal, 12(1), 34-58.

Zadeh, A. H., Zolbanin, H.M., Sengupta, A., & Schultz, T. (2020). Enhancing ERP Learning Outcomes through Microsoft Dynamics. Journal of Information Systems Education, 31(2), 83-95.

APPENDIX

Student Handout

-----------Beginning of student handout----------------------------

You have recently been hired by an organization that manufactures bicycles and sells these to wholesalers across the country. Due to disruptions in the supply-chain there is a shortage of bicycles. This shortage will last a few months. There are confirmed and unconfirmed orders (additional requests). As there are not enough bicycles to be distributed you will prioritize based on the strategies given in Table 1. Assume that you have already categorized your customers into the following confirmation strategy types.

Table 1: Confirmation Strategies

CONFIRMATION STRATEGY	DETAILS
WIN	Confirm all orders whether fully confirmed or not confirmed. In strategic terms, these customers are very important.
GAIN	Confirmed fully confirmed, for unconfirmed - if possible- improve if stocks are available.
REDISTRIBUTE	May lose or gain both confirmed or unconfirmed.
FILL	May Lose confirmation but will not gain anything (will lose unconfirmed)
LOSE	Will lose both confirmed and unconfirmed quantity - In strategic terms, these customers may not be important.

Question 1:

Four examples of recent item shortages.

Example	Details (around 10 words)	Year
1		
2		
3		
4		

Question 2: Why would a company prioritize and rank its customers (rank). Provide 4 reasons.

Four reasons a company would prioritize customers.	
Reason	Details (around 10 words)
1	
2	
3	
4	

Question 3: For Situations 1 through 5, fill the last column based in strategies provided in table 2. Assume total available units in each case includes all items available (this includes all confirmed and unconfirmed orders that have not been shipped).

Scenario 1: Total available 51000 units.

Confirmation Strategy	Initial Confirmation	Additional Request	Total units Given
WIN	10000	2000	
GAIN	10000	2000	
REDISTRIBUTE	10000	5000	
FILL	10000	2000	
LOSE	10000	0	

Scenario 2: Total available 29000 units

Confirmation Strategy	Initial Confirmation	Additional Request	Total units Given
WIN	10000	0	
GAIN	10000	2000	
REDISTRIBUTE	0	0	
FILL	10000	2000	
LOSE	10000	0	

Scenario 3: Total available 25000 units.

Confirmation Strategy	Initial Confirmation	Additional Request	Total units Given
WIN	0	2000	
GAIN	0	2000	
REDISTRIBUTE	15000	2000	
FILL	10000	2000	
LOSE	10000	0	

Scenario 4: Total available 25000 units.

Confirmation Strategy	Initial Confirmation	Additional Request	Total units Given
WIN	0	2000	
GAIN	0	2000	
REDISTRIBUTE	0	0	
FILL	1000	2000	
LOSE	30000	0	

Scenario 5: Total available 29000 units.

Confirmation Strategy	Initial Confirmation	Additional Request	Total units Given
WIN	0	0	
GAIN	0	2000	
REDISTRIBUTE	10000	5000	
FILL	10000	2000	
LOSE	0	0	

-----------End of student handout----------------------------

Solution

Answers to questions 1 could include examples of recent item shortages are shown in table 2.

Answer to question 2: why a company would prioritize and rank its customers (rank).

Rationality of Backorder processing

- The power of the partner- Powerful sellers may expect that their customers are willing to wait for their product rather than switch to a competitor. On the other side, powerful buyers are more likely to switch to competitors.
- Problem customers. Perhaps such customers in this group are blocked for bad credit and are assigned to lower category.
- Customers with history of relationship with the company (Loyal Customers) may be given high priority
- Starting a new relationship with potential powerful customer will lead to higher priority.
- At times, the company is bound by contracts with customers to honor all orders. Such partners will be put in the WIN list as pre-existing contracts cannot be reneged
- Priority could be mandated too. For example, during the baby formula shortage of 2022, priority was given to hospitals.

Answer to Question 3

Note: To simplify the process, the cells colors can be colored based on decisions to be made. Green meaning- must send material; Blue means- depends on availability and Red means – cancel/postpone orders.

Scenario 1: Total available units 510000

	Initial Confirmation	Additional request	Total units Given
Win	10000	2000	12000
Gain	10000	2000	12000
Redistribute	10000	5000	15000
Fill	10000	2000	10000
Lose	10000	0	0

Scenario 2: Total available units 29000

	Initial Confirmation	Additional request	Total units Given
Win	10000	0	10000
Gain	10000	2000	12000
Redistribute	0	0	0
Fill	10000	2000	7000
Lose	10000	0	0

Scenario 3: Total available units 25000

	Initial Confirmation	Additional request	Total units Given
Win	0	2000	2000
Gain	0	2000	2000
Redistribute	15000	2000	17000
Fill	10000	2000	4000
Lose	10000	0	0

Scenario 4: Total available units 25000

	Initial Confirmation	Additional request	Total units Given
Win	0	2000	2000
Gain	0	2000	2000
Redistribute	0	0	0
Fill	1000	2000	1000
Lose	30000	0	0

Scenario 5: Total available units 29000

	Initial Confirmation	Additional request	Total units Given
Win	0	0	0
Gain	0	2000	2000
Redistribute	10000	5000	15000
Fill	10000	2000	10000
Lose	0	0	0

Explanation of scenarios

Scenario 1: WIN gets first. FILL does not get any additional requests. LOSE gets confirmed and unconfirmed requests rejected.

Scenario 2: FILL does not get any additional request. LOSE gets confirmed and unconfirmed requests rejected.

Scenario 3: WIN gets additional request confirmed. LOSE gets confirmed and unconfirmed requests rejected.

Scenario 4: Even when items are available, items are not given to LOSE. LOSE gets confirmed and unconfirmed requests rejected. FILL too does not get unconfirmed orders.

Scenario 5: FILL group does not get any additional request. LOSE gets confirmed and unconfirmed requests rejected.

Spreadsheet Model using scenario 1 with formulas.

	A	B	C	D	E
1		Initial Confirmation	Additional request	Given	Total Remaining
2					51,000
3	Win	10000	2000	=IF(E2>B3+C3,B3+C3,E2)	=IF(E2>B3,E2-B3-C3,0)
4	Gain	10000	2000	=IF(E3>B4+C4,B4+C4,E3)	=IF(E3>B4,E3-B4-C4,0)
5	Redistribute	10000	5000	=IF(E4>B5+C5,B5+C5,E4)	=IF(E4>B5,E4-B5-C5,0)
6	Fill	10000	2000	=IF(E5>B6,B6,E5)	=IF(E5>B6,E5-B6-C6,0)
7	Lose	10000	0	0	=IF(E6>B7,E6-B7-C7,0)
8					

DECISION SCIENCES INSTITUTE

Multi-Period Pricing and Product Improvement:
How Informative Are Product Ratings to New Customers?

Alireza Yazdani
California State Polytechnic University, Pomona
Email: ayazdani@cpp.edu

Hossein Rikhtehgar Berenji
Pacific University
Email: hossein@pacificu.edu

ABSTRACT

Employing a multi-period game-theoretic setting, we analyze a forward-looking firm's pricing and product improvement strategies, facing multiple generations of customers who potentially receive a performance signal from prior product ratings. We show that this signal is fully informative in equilibrium only if the product has received a mix of positive and negative reviews. We further establish that the firm increases its optimal price after a period wherein the true performance proves to be greater or slightly less than the ex-ante expectation, and decreases the price otherwise.

KEYWORDS: Pricing, Game Theory, Customer Ratings

DECISION SCIENCES INSTITUTE

Obsessive-Compulsive Branded Apparel Buying Behavior in Kuwait: The Mediating Role of Brand Attachment

Mohamed M. Mostafa
Gulf University for Science and Technology, Kuwait
Email: mostafa@usa.com

ABSTRACT

Deeply rooted in the fields of medicine and psychiatry, obsessive-compulsive behavior continues to gain considerable attention in marketing. This study aims to examine the major factors driving obsessive-compulsive buying behavior and the mediating role of brand attachment in Kuwait. The study used partial least squares structural equation modeling (PLS-SEM) on a random sample of 302 branded apparel consumers. The results show that materialism and utilitarian values are positively and significantly related to brand attachment, which, in turn, mediates the relationship between materialism, utilitarian and hedonic values, and obsessive-compulsive buying behavior. The study contributes to the literature by investigating the predictors of obsessive-compulsive buying behavior in a non-Western collectivist culture. Furthermore, the study highlights the essential role played by brand attachment as a mediator within the context of branded apparel.

KEYWORDS: Obsessive-Compulsive Buying Behavior, Brand Attachment, Survey Research, Kuwait, Partial Least Squares

DECISION SCIENCES INSTITUTE
Option Pricing Simplified

Nilanjana Chakraborty
Freelance Researcher, India.
dr.nila.chakraborty@gmail.com

Mohammed Elgammal
Qatar University
m.elammal@qu.edu.qa

ABSTRACT

Here we present three option-pricing models, one based on the underlying asset and two based on indices that are averages of the prices of options having same particulars but different strike prices. These models estimate an option price directly from the market movements in the prices of the underlying asset or the options having different strike prices. Besides their ease of computation, the average pricing errors of these new models are comparable to or less than those of the extant models like the Black-Scholes model, the Heston-Nandi GARCH(1,1) model and the Conditional Black-Scholes model, making the former preferable.

KEYWORDS: Option pricing, Black-Scholes Model, Heston-Nandi GARCH Model, Conditional Black-Scholes Model, Index Models, *JEL Codes:* G13

DECISION SCIENCES INSTITUTE

Optimizing Supply Chain Collaborations in the Face of Uncertainty and Asymmetric Information

Chuqiao Peng
University of Oklahoma
Email: chuqiao@ou.edu

Desmond (Ho-Fu) Lo
Santa Clara University
Email: hlo@scu.edu

Kenneth J. Petersen
University of Oklahoma
Email: petersen@ou.edu

Qiong Wang
University of Oklahoma
Email: qiongwang@ou.edu

ABSTRACT

We investigate how collaboration strategies respond to uncertainty and asymmetric information and hence achieve resilient supply chains. Building on information processing theory, we find that uncertainty positively affects buffering and bridging while asymmetric information negatively affects buffering and bridging. More interestingly, asymmetric information impairs bridging more than it does on buffering. In practice, firms may adopt both bridging and buffering concurrently. Nevertheless, their respective contributions differ. Buffering strategies enhance the supply chain's resilience more effectively compared to bridging.

KEYWORDS: Uncertainty, Asymmetric information, Buffering, Bridging

DECISION SCIENCES INSTITUTE

Patterns of Gender-Specific Multiple Chronic Conditions in Working-Age Adults, their Generational Disparity and Community-level Socioeconomic Determinants

Ajit Appari
Boston University
Email: aappari.work@gmail.com

Maria Ukhanova
University of Texas Health Science Center at Houston
Email: maria.ukhanova.1@uth.tmc.edu

ABSTRACT

Multiple chronic conditions (MCC) are a major challenge among working-age Americans. We present research that identifies MCC patterns using exploratory factor analysis on six-years (2008-2013) claims history of 452,834 commercially insured adults from three birth-cohorts (Generation-X/Millennials; younger, and older -baby boomers) in Texas. Our analysis yielded eight gender-specific MCC patterns: (men) cardiovascular-metabolic, metabolic-urologic, hepatorenal-autoimmune, musculoskeletal; (women) metabolic, cardiovascular, renal-autoimmune, musculoskeletal. Next, we evaluated the risk of developing these MCC patterns across generations and by (zip-code) community level socioeconomic determinants using multilevel mixed-effects logistic model. Across all patterns, we found higher risk among baby boomers and individuals living in minority-dominant areas.

KEYWORDS: Multiple Chronic Conditions, Working-Age Americans, Multilevel Logistic Regression, Exploratory Factor Analysis, Social Determinants

INTRODUCTION

Multiple chronic conditions (MCC), the co-existence of two or more chronic conditions, has become a major public health issue in the United States (DHHS 2010) and requires renewed focus to advance patient-centered care for people with MCC (Bierman and Khanna 2019; Bierman et al. 2021). Approximately 25 percent of adult Americans are living with MCC (Ward et al. 2014), and about two out of three older Americans live with MCC (Marengoni et al. 2011). Moreover, the absolute number of individuals with MCC is higher among working-age adults (Barnett et al. 2012; Machlin and Soni 2013; Rocca et al. 2014). Although, a significant body of MCC research has emerged in the last two decades, our understanding of MCC patterns for American population remains limited. Several systematic reviews acknowledged multiple approaches in vogue for identifying and describing MCC (Fortin et al. 2012; Huntley et al. 2012; Violan et al. 2014). Most studies used count or frequently occurring tuples of chronic conditions to identify MCC (Ford et al. 2013; Freund et al. 2012; Machlin and Soni 2013; Rocca et al. 2014; Roberts et al. 2015; Violan et al. 2014). However, these approaches are inadequate for effectively managing individuals' or population health. An emerging body of research sought to address this gap by identifying MCC patterns using advanced methods including cluster analysis, exploratory factor analysis, and agglomerative hierarchical clustering (e.g., Abad-Díez et al. 2014; Marengoni et al. 2010; Prados-Torres et al. 2012; Schäfer et al. 2014; van den Bussche et al. 2011; Violan et al. 2013). Surprisingly, most of this literature focused on non-US populations, with the exception of few studies (e.g., Cornell et al. 2007; Newcomer et al. 2011) that examined Medicare beneficiaries, veterans or patients of a single healthcare provider in the US. Furthermore, there is a growing recognition for the need of gender-specific approaches to medical diagnosis and treatment (Glezerman 2016). Finally, a significant body of research at the intersection of health services research and medical sociology has studied small-area effects on health and health care delivery for patients with specific chronic conditions (Badland et al. 2013; Diez Roux 2010); however, to best of our knowledge, no study characterized MCC patterns across small-area communities.

This study seeks to address two important research gaps in the MCC literature: identification of gender-specific MCC patterns for working-age American adults based on the correlational structure of chronic conditions; and evaluating the variation in risk of developing these gender-specific MCC patterns across communities characterized by socioeconomic determinants. For the first aim, we employed exploratory factor analysis on a six years (2008-2013) long-run claims history of working-age adults residing in Texas and covered continuously by a commercial carrier. Next, for the second aim, we separately estimated mixed effect logistic regression models for each gender-specific pattern to evaluate how the risk of developing them varied across three birth-cohorts (younger baby boomers and older baby boomers compared to Generation-X/Millennials) to capture generational effects, and across the distributional spectrum of ZCTA (ZIP-code tabulation area) level socioeconomic determinants to explain community-level socioeconomic disparity.

Our study makes two important contributions to literature on multiple chronic conditions. First, this study is one of the first to identify gender-specific patterns of multiple chronic conditions for a large statewide working-age American adults using exploratory factor analysis. Next, this study identifies community-level disparity in how the propensity of developing these patterns varies across small-areas having different socioeconomic positions (e.g. minority concentration, rurality, household median income, educational attainment, and foreign-born population concentration).

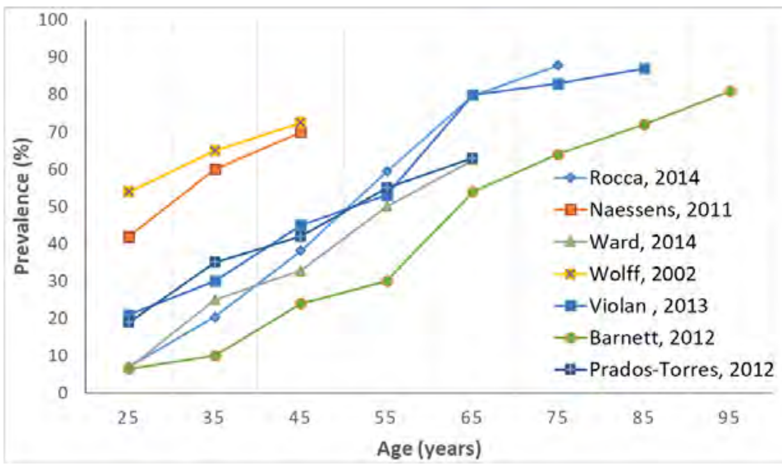
LITERATURE REVIEW

The rising prevalence of multiple chronic conditions is a major public health issue, particularly among working-age adults. Approximately 25 percent of Americans are living with two or more chronic conditions (Ward et al., 2014) and this number increases to two people out of three, specifically in the population older than 65 (Marengoni et al., 2011). The prevalence of MCC often increases steeply with age (see Figure 1). The prevalence of MCC exponentially rises from 7% to 87% across 10 years segments in the age ranges 20-80plus years (Rocca et al., 2014), about 40% to 67% in the age range 20 – 45 years for adults (Naessens et al., 2011; Wolff 2002).

Several analytical approaches have been used in the literature to study MCC among individuals in different contexts including patient populations affiliated to a health system, residents of designated geographic area [e.g., city, county, state, and nation], elder population (>=65 years), younger population (typically 18-64 years), and by gender segment (Huntley et. al. 2012; Suls et al. 2021). One set of studies focused on the determination of MCC prevalence based on count of chronic conditions, tuples of most frequently co-occurring chronic conditions, and clustering of chronic conditions, e.g., hierarchical clustering, network analysis, exploratory factor analysis (Violan el al., 2013; Prados-Torrse et al., 2014; Rocca et al., 2014; Jackson et al., 2016).

Further, such studies often examined how MCC prevalence varies in terms of explanatory factors such as age, gender, and social determinants. Another set of studies evaluated MCC using weighted index such as Charlson comorbidity index, Elixhauser index (Charlson et al. 1987; Elixhauser et al. 1998; Klabunde et al. 2007; Romano, Roos & Jollis 1993; Sundarrajan et al. 2004; van Walraven et al. 2007). These estimates of MCC are often used either as confounding factor or explanatory factors in studying of select anchor chronic conditions such as diabetes, asthma, mental health (Kilbourne et al., 2004; Piette et al., 2006; Bhattacharya et al., 2014) or as one of the explanatory factors in cost and utilization of healthcare services (Wolff et al., 2002; Glynn et al., 2011; Naessens et. al., 2011).

Figure 1: Trend of MCC Prevalence by Age Reported in the Select Studies.



The measurement of MCC prevalence reported in past research could potentially be influenced by factors that characterize study design itself besides the clinically relevant factors (Fortin et al., 2012). For example, many studies used baseline set of chronic diseases that varies in counts from 10 (Ward et al., 2014) to as large as 185 (Kadam et al., 2007). Similarly, factors such as study sample (primary care population based, hospital admissions, physician's office visits), and data collection methods (surveys, medical records, administrative databases) contributed to variations in MCC prevalence and made it challenging to estimate prevalence (Fortin et al., 2012; Boyd & Fortin 2010; Violan et al. 2014; Fortin et al., 2010; McGrail et al., 2016). Besides characteristics of study methodology, other factors may affect the variation in MCC prevalence.

Social determinants of health, such as age, gender, and socioeconomic status affect the emergence and evolution of chronic conditions; however, they are poorly understood in the context of multimorbidity (Álvarez-Gálvez et al. 2023). Typically, MCC prevalence is estimated to be lower in younger population compared to older population but in absolute number younger individuals with MCC outnumber the older people. For example, Rocca and colleagues found that the prevalence of MCC in working age adults (64 and younger was only 14.4%; while in the population older than 65, prevalence increased to 77.3%, however, the absolute number for the first group was 17,838 individuals, while the 65 and older group lowered to 13,606 individuals (Rocca et al., 2014). Similar observations in a larger sample were reported by Barnett et al.

(2012) and based on a Scottish population where 210,500 individuals with MCC were younger than 65, as compared to 194,966 individuals older than 65 with trends in prevalence close to the US study (Barnett et al., 2012). In the meta-analysis regarding the prevalence of MCC and MCC clusters, women suffer more from MCC than men (Violan et al. 2014; Agur 2016). There is a difference in prevalence and type of MCC between genders by age group (Agur 2016). People with a lower social-economic status and patients with mental health disorders also have higher odds of developing MCC (Marengoni et al. 2011; Violan et al. 2014; Barnett et al. 2012; Roberts et al., 2015). The black population has higher rates of MCC as compared to white and Hispanic populations (Ward et al., 2014; Rocca et al., 2014). People in deprived areas develop MCC 10-15 years earlier than those in higher socio-economic areas, especially if one of the conditions is mental disorders (Barnett et al., 2012).

The patterns of chronic conditions are different between age groups. Older patients tend to have more MCC and lower functional status, while the younger population tends to have combinations of physical and mental disorders (Barnett et al. 2012). Knowing the prevalence of MCC and associated factors helps to understand the epidemiology of MCC and help to effectively organize healthcare (Barnett et al. 2012).

Specific conditions tend to occur in "clusters" - groups of certain chronic conditions that appear together in the same patient at "greater rates than would be expected by chance alone" (Marengoni et al. 2011). A simple count of the number of diseases does not provide insight into nonrandom combinations of MCC, in contrast with advanced statistical methods (Holden et al. 2011). The number of MCC clusters varies between 2 and more than 50 (Table 1).

Table 1. Patterns of multiple chronic conditions and statistical methods in selected studies§

Study	Population Studied	Clustering Method	#Clusters	MCC Cluster Details
García-Olmos et al., (2012)	# Individual: 198,670 Type: Outpatients, 15+ years # Baseline diseases: 26 Country: Spain	Multiple correspondence analysis	4	Pattern A: hypertension, disorder of lipid metabolism, diabetes 2, cardiac arrhythmia. Pattern B: cerebrovascular disease, ischemic heart disease (exl AMI), chronic renal failure, congestive heart failure. Pattern C: anxiety/depression, thyroid disease, asthma, schizophrenia and effective psychoses. Pattern D: obesity, osteoporosis, hearing loss, malignant neoplasms, degenerative joint disease, benign prostatic hypertrophy, emphysema, chronic bronchitis, and COPD.
Holden, et al. (2011)	# Individual: 78,430 Type: Employees, 18-70 years # Baseline diseases: 23 Country: Australia	Exploratory factor analysis with tetrachoric correlation	6	Factor1: arthritis, osteoporosis, other chronic pain, bladder problems, irritable bowel; Factor2: asthma, COPD, allergies; Factor3: back/neck pain, migraine, other chronic pain, arthritis; Factor4: high blood pressure, high cholesterol, obesity, diabetes, fatigue; Factor5: CVD, diabetes, fatigue, high blood pressure, high cholesterol, arthritis; Factor6: irritable bowel, ulcer, heartburn, other chronic pain.

Jackson, et al., (2016)	# Individual: 13,715 Type: population based study on women's health, females 45-50 years # Baseline diseases: 31 Country: Australia	Exploratory factor analysis with tetrachoric correlation, and principal factor	5	Psychosomatic: anxiety, depression, severe tiredness, poor memory, severe headache, chest pain, vision problems, bowel problems, palpitations, joint pain. Musculoskeletal: joint pain, other arthritis, osteoarthritis, back pain, rheumatoid arthritis, osteoporosis. Cardiometabolic: rheumatoid arthritis, impaired glucose tolerance, diabetes, stroke, chronic fatigue syndrome, heart disease, hypertension, other psychiatric disorders. Cancer: other cancer, breast cancer, cervical cancer. Respiratory: asthma, bronchitis/emphysema, breathing difficulty, allergies.
Kirchberger et al. (2012)	# Individual: 4,127 Type: General population, 65-94 years # Baseline diseases: 13 Country: Germany	Exploratory factor analysis with tetrachoric correlation, and principal factor method	4	Cardiovascular/metabolic disorders, Liver/lung/eye disorders. Mental/neurologic disorders. Gastrointestinal disorders and cancer.
Marengoni, et al. (2010)	# Individual: 1,332 Type: inpatient, 65+years # Baseline diseases: 19 Country: Italy	Cluster analysis with average linkage	8	1. Liver cirrhosis, malignancy. 2. Chronic Renal Failure, anemia. 3. Hypertension, dyslipidemia. 4. Gastric diseases, intestinal diseases. 5. Dementia, arthritis 6. Thyroid dysfunction, anxiety. 7. Diabetes mellitus, CHD, CVD. 8. HF, AF, COPD, prostate hypertrophy.
Newcomer, et al. (2011)	# Individual: 15,480 Type: Outpatient, inpatient, and cancer registry, Adults 21+ years # Baseline diseases: 17 Country: USA	Agglomerative hierarchical cluster analysis	10	1. Chronic pain with mental health conditions. 2. Diabetes with obesity and mental health conditions. 3. Kidney disease with diabetes and obesity. 4. Mental health conditions and obesity in younger adults. 5. Frailty in the elderly: mental health conditions, diabetes, obesity, stroke, cardiac disease, kidney disease, skin ulcers, dementia. 6. Cardiac disease and obesity, diabetes. 7. Chronic obstructive pulmonary disease (COPD) with obesity and mental health conditions 8. Gastrointestinal bleeding with obesity and mental health conditions. 9. Abdominal and orthopedic surgeries with obesity. 10. Cancer with obesity and mental health conditions.
Prados-Torres et al. (2012)	# Individual: 275,682 Type: primary care patients, 15+ years # Baseline diseases: 13 to 39 by age groups & gender Country: Spain	Exploratory factor analysis with tetrachoric correlation, and principal factor method	5	1. Cardio-metabolic: diabetes, hypertension, obesity and dyslipidemia. 2. Psychiatric-substance abuse: psychosis and neurosis. 3. Mechanical: mechanical-obesity-thyroidal. 4. Psychogeriatric: dementia, behavioral problems, Parkinson's disease, osteoporosis, chronic skin ulcers and iron deficiency. 5. Depressive: depression and behavioral disorders.

Schäfer et. al. (2014)	# Individual: 149,280 Type: Outpatient, 65+ years # Baseline diseases: 46 Country: Germany	Network analysis	2	1. Cardiovascular and metabolic disorders. 2. Anxiety, depression, somatoform disorders and pain
Van den Bussche, et al. (2011)	# Individual: 123,224 Type: Inpatients, any age # Baseline diseases: 46 Country: Germany	Tuples of disease, Conditional probability	50	1. Hypertension, lipid metabolism disorders, chronic low back pain. 2. Hypertension, chronic low back pain, osteoarthritis. 3. Hypertension, lipid metabolism disorders, chronic ischemic heart disease. 4. Hypertension, lipid metabolism disorders, diabetes mellitus. 5. Hypertension, lipid metabolism disorders, osteoarthritis. 6. Lipid metabolism disorders, chronic low back pain, osteoarthritis. 7. Hypertension, lipid metabolism disorders, purine/pyrimidine metabolism disorders/gout. 8. Hypertension, chronic low back pain, chronic ischemic heart disease. 9. Hypertension, chronic low back pain, diabetes mellitus. 10. Hypertension, diabetes mellitus, chronic ischemic heart disease.
Violan, et al., (2013)	# Individual: 18,126 Type: population based study, primary care, 15+ years # Baseline diseases: 27 Country: Spain	Pairs and tuples of disease, Conditional probability	351 pairs, 2925 triplets	1.Depression or anxiety and mental disorder. 2. Hypertension and diabetes mellitus. 3. Hypertension and prostatic disorder. 4. Hypertension and osteoarthritis or rheumatism. Etc.

In recent years, new studies have proposed constructing clusters of MCC using exploratory factor analysis (Kubinger 2003; Calderon-Larranaga et al., 2013; Kirchberger et al., 2012; Prados-Torres et al., 2012; Schäfer et al., 2010, Holden et al., 2011). In addition to exploratory factor analysis, Prados-Torres et al. (2014) in the systematic review of MCC patterns, identified such methods as agglomerative hierarchical clustering (Cornell et al., 2007; Goldstein et al., 2008; Newcomer et al., 2011), and observed-to-expected ratio (Van den Bussche et al., 2011; Freund et al., 2012; Wong et al., 2011) to identify disease clusters. There is a need for more research using large population-based data to assess the complexity of multiple chronic conditions (Holden et al., 2011). Administrative claims data is beneficial for this kind of analysis.

In a recent systematic review of literature, Violan et al. (2014) identified the prevalence of several clusters of MCC, in combinations of two or three conditions. The most frequent combinations of chronic conditions in all age groups were those that are related to osteoarthritis and cardio-metabolic clusters. For the population younger than 65, more typical clusters were cardio-metabolic and mechanical-obesity-thyroidal (Britt et al., 2008; Barnett et al., 2012; Violan et al., 2013; Brilleman et al., 2013). In the population older than 65, more common clusters were cardio-metabolic and hypertension-osteoarthritis (Lochner et al., 2013; Marengoni et al., 2009; Cornell et al., 2007; Newcomer et al., 2011).

Building upon this emerging stream of research, we examine patterns of multiple chronic conditions among working age adults using exploratory factors analysis separately for men and women. Subsequently, we evaluate how the risk of developing such MCC pattern differs across birth cohort, and across a set of social determinants.

METHODS

Data

This retrospective study included a cohort of 452,834 working-age adults (18-64 years old; men=202,776; women=250,058), continuously enrolled in preferred provider organizations (PPO) or PPO+ plans during 2008-2013. Individuals were allowed 90-days gap between consecutive enrollments. Each individual had at least two outpatients and one inpatient claims to ensure reliable identification of chronic conditions (Schultz et al. 2013; Wolff et al. 2002). For ZCTA-level socioeconomic factors, we used the 2008-2012 American Community Survey 5-Year Estimates. The institutional review board at the University of Texas Health Science Center at Houston approved this study.

Measurement of Chronic Conditions

We used AHRQ clinical classification software and 62 chronic condition category indicators, based on modified version of chronic condition classification map proposed by Magnan (2015). The Magnan's (2015) classification system maps 4,132 diagnosis codes (international classification of diseases, 9th revision) into clinically relevant chronic condition categories. We excluded mental health diagnoses or sexually transmitted infections related categories. Individuals were categorized for chronic conditions if underlying diagnosis codes appeared in at least one inpatient claim and/or two outpatient claims with different service dates (Elixhauser et al. 2016; Lochner and Cox 2013; Rocca et al. 2014).

Measurement of Individual-Level Covariates

All individuals were categorized into three birth-cohorts using the effective date of first enrollment in 2008 or before (baseline year of study period) as reference start date: (1) Generation-X/Millennials [reference group] born during 1965-1990, aged 18-44 years; (2) Younger Baby Boomers born during 1957-1964, aged 45-54 years; and (3) Older Baby Boomers born during 1943-1956, aged 55-64 years. Although, Generation-X (born during 1965-1980) and Millennials (1981-1999) cohorts are considered separately in social science research, we combined them to create the younger group of working-age adults, 18-44 years old.

Other individual-level covariates include insurance subscription type (primary vs. dependent), employer's industry (two-digit standard industry classification: extended to both subscriber types), and additional chronic disease burden (conditions not utilized for allocating individuals to all MCC patterns).

Measurement of Community (ZCTA) Level Covariates

To investigate area-level disparity in the risk of developing MCC patterns, we used several ZCTA-level socioeconomic factors, specifically minority concentration, educational attainment, annual median household income, rurality, and foreign-born population concentration. These factors have been used in research to characterize socioeconomic status of the place of residence to elicit geographic disparity (Abbass et al. 2017; Basu, Avila, and Ricciardi 2016; Diez Roux 2010; Herrin et al. 2015).

For minority concentrations, ZCTAs were classified using three indicators representing highest percentage of race/ethnicity: non-Hispanic Blacks, Hispanics, and non-Hispanic Whites [majority reference]. For income, five indicators representing the quintiles of annual median household income were used. For rurality, three indicators as urban [reference], large rural, and small rural areas were used. For remaining covariates, we constructed three indicators representing distribution: bottom (below 25th percentile), middle (25th-to-75th percentile: reference), and top (above 75th percentile) quartiles.

Statistical Analysis
Identification of MCC Patterns: We extracted gender-specific MCC patterns using Exploratory Factor Analysis (EFA) on a baseline set of 42 chronic conditions eliciting their correlational structure (Pett, Lackey and Sullivan 2003; Kline, 2014). EFA has been applied in uncovering MCC patterns from claims data on European population (e.g., Abad-Díez et al. 2014; Holden et al. 2011; Newcomer et al. 2011; Prados-Torres et al. 2012; Schäfer et al. 2014). The baseline set was selected by including conditions with one percent or more prevalence in at least one birth-cohort [See online Appendix-C]. To extract clinically relevant factors from dichotomously coded data, we estimated a tetrachoric correlation matrix and subsequently applied principal factor method using the *factormat* command and varimax rotation (Kirchberger et al. 2012, Jackson et al. 2016). We applied additional criteria to finalize gender-specific MCC patterns: the scree plot of factors, and eigenvalues above one (Holden et al. 2011); at least two conditions load meaningfully in each factor; the cumulative proportion of variance explained by target factors remains about 80 percent (Prados-Torres et al. 2012); parallel analysis (Fujimoto et al. 2014); and the sampling adequacy measure Kaiser-Meyer-Olkin index above 0.5 (Kaiser 1974). Finally, the membership of conditions to each pattern was determined using 0.5 as threshold for factor loading (Hallman et al. 2003), and uniqueness statistic below 0.6 was used for conditions with near threshold factor loadings (Costello and Osborne, 2005). Individuals were assigned to each MCC pattern if they had at least two member conditions for the respective pattern.

Evaluation of Risk of Developing MCC Patterns. We, first, described each gender-specific MCC pattern by estimating: (a) prevalence of patterns and their member conditions by birth-cohort to describe aging trends, and (b) prevalence of MCC pattern combinations with respect to the full sample and within the sub-population of individuals with 2plus chronic conditions; and (c) pattern prevalence with respect to individual factors and ZCTA-level socioeconomic factors.

Next, we estimated multilevel mixed effects logistic regression model separately for each gender-specific pattern with ZCTA-level random effects to evaluate the association of MCC patterns with an individual's demographic characteristics and their residential ZCTA-level socioeconomic factors. Raw claims data was processed in SAS and analyses were performed in Stata13.0 (StataCorp).

RESULTS

The analytic sample included 452,834 working-age adults (men=202,776[44.8 percent]; women= 250,058[55.2 percent]) enrolled in PPO/PPO+ plans during the study period, 2008-2013, and had at least two outpatients and one inpatient claims. Table 1 shows summary statistics for the study sample and all enrollees in the baseline year 2008 with respect to individual and community factors. In the study sample, the average age was 42.8 years (standard deviation 10.5 years), 47.2 percent were Generation-X/Millennials, 28.1 percent were younger baby boomers, and 24.7 percent were older baby boomers. The ZCTA-level residential distributions include 23.6 percent in Hispanic dominant areas, 3.4 percent in non-Hispanic Black dominant areas, 76.8 percent in urbanized areas, 39.2 percent in high-income (median household income $60,000 plus), and 10.3 percent in low income (median household income below $35,000) areas.

Table 1: Characteristics of Study Sample (continuously enrolled 2008-2013) and All Commercially Insured Working-Age Adult Texas Residents in the Baseline Year-2008

CHARACTERISTICS	STUDY SAMPLE, 2008-13 COUNT(%)	TOTAL ENROLLMENT, 2008 COUNT (%)
Total Individuals Count	452,834(100.0)	1,157,613(100.0)
Individual Level Factors		
Age, Years‡	42.8(10.5)	42.2(12.3)
Birth Cohort‡		
Generation-X/Millennials	213,610(47.2)	583,801(50.4)
Younger Baby Boomers	127,368(28.1)	258,194(22.3)
Older Baby Boomers	111,856(24.7)	315,618(27.3)
Gender		
Male	202,776(44.8)	545,141(47.1)
Female	250,058(55.2)	612,472(52.9)
Insurance Policy Holder		
Main subscriber	342,028(75.5)	841,460(72.7)
Dependent/ Others	110,806(24.5)	316,153(27.3)
Industry Group		
Construction	10,260(2.3)	36,194(3.1)
Finance/Insurance/Real Estate	45,483(10.0)	99,546(8.6)
Manufacturing	49,415(10.9)	138,276(11.9)
Mining, Agriculture	16,977(3.7)	54,811(4.7)
Retail Trade	16,521(3.6)	105,857(9.1)
Services	187,243(41.3)	448,686(38.8)
Trans., Comm., Electricity/ Gas, & Sanitary Services	17,016(3.8)	43,843(3.8)
Unknown	70,252(15.5)	143,811(12.4)
Wholesale Trade	39,667(8.8)	86,589(7.5)
Community Area (ZCTA) Level Socioeconomic Factors		
Dominated Race/Ethnicity		
Non-Hispanic Whites	330,480(73.0)	816,523(70.5)
Non-Hispanic Blacks	15,518(3.4)	43,470(3.8)
Hispanics	106,836(23.6)	297,620(25.7)
Rurality of Area		
Urban	347,576(76.8)	903,560(78.1)

Large rural	52,914(11.7)	130,233(11.3)
Small rural	52,344(11.6)	123,820(10.7)
Median Household Income		
Quintile-1	50,156(11.1)	138,379(12.0)
Quintile-2	87,052(19.2)	226,717(19.6)
Quintile-3	77,149(17)	198,930(17.2)
Quintile-4	108,986(24.1)	277,805(24.0)
Quintile-5	129,491(28.6)	315,782(27.3)
Education- Population Concentration with Bachelor or Above Degree(% of population)		
Lower Quartile	88,380(19.5)	237,814(20.5)
Middle 50th	232,633(51.4)	592,747(51.2)
Top Quartile	131,821(29.1)	327,052(28.3)
Foreign-born Population Concentration (% of population)		
Lower Quartile	164,751(36.4)	400,136(34.6)
Middle 50th	224,007(49.5)	575,741(49.7)
Top Quartile	64,076(14.1)	181,736(15.7)
‡ All working age adults were categorized into their birth cohort based on their age as on Jan-01, 2008: (a) Generation-X/Millennials of 18-44 years old; (b) Younger Baby Boomers of 45 to 54 years old; and (c) Older baby Boomers of 55 to 64 years old.		

Identification of Multiple Chronic Condition Patterns

The exploratory factor analysis revealed four MCC patterns for each gender. Table 2 shows factor loadings and uniqueness statistics of relevant chronic conditions on each MCC pattern. For working-age men, we identified cardiovascular-metabolic [factor loadings λ's: 0.53–0.87; member conditions=10], hepatorenal-autoimmune [λs: 0.53–0.73; member conditions=6], metabolic-urological [λ's: 0.50–0.69; member conditions=6], and musculoskeletal [λ's: 0.50–0.61; member conditions=6] patterns. For working-age women, we identified cardiovascular diseases [λ's: 0.53–0.80 for nine member conditions], musculoskeletal [λ's: 0.52–0.59 for six member conditions], metabolic [λ's: 0.55–0.69 for four member conditions], and renal-autoimmune [λ's: 0.48–0.65 for four member conditions] patterns. The cardiovascular based pattern for both genders had nine common chronic conditions including hypertension, coronary atherosclerosis, acute myocardial infarction, congestive heart failure, cardiomyopathy and structural heart disease, heart valve disorder, conduction disorder or cardiac dysrhythmia, peripheral atherosclerosis, and cerebrovascular disease; with hyperlipidemia as an additional condition for men. The metabolic based pattern for both gender had three common conditions: hypertension, hyperlipidemia, and diabetes mellitus; obesity occurred as an additional condition for women only, and men had three additional conditions: other endocrine, benign prostatic hypertrophy (BPH), and male genitourinary disorders excluding BPH. The musculoskeletal pattern for both genders had identical set of six conditions: osteoarthritis, rheumatoid arthritis, other central and peripheral nervous system disorders, lupus, back problems, and other musculoskeletal including osteoporosis. Finally, the renal-autoimmune based pattern in both gender had three common conditions: chronic renal failure, anemia and other non-cancer hematoma disorder, and immunity disorder; lupus occurred as additional condition for women, and men had three additional conditions including chronic hepatitis, kidney and vesicoureteral disease (excluding renal failure), and chronic liver disease (excluding chronic hepatitis).

Table 2: Identification of Gender-Specific Multiple Chronic Condition (MCC) Patterns Using Exploratory Factor Analysis on the Diagnoses History of 452,834 Working-Age Adults During 2008-2013 in Texas §

Chronic Conditions and Statistics	**MCC Patterns: Men (n=202,776)**						**MCC Patterns: Women (n=250,058)**				
	CVMB	**HRAI**	**MTBU**	**MSK**	**Ψ***		**CVD**	**RAI**	**MTB**	**MSK**	**Ψ***
Eigenvalues	12.08	2.39	1.60	1.35			11.20	1.19	1.49	2.32	
Variance explained	0.56	0.11	0.07	0.06			0.55	0.06	0.07	0.11	
Coronary Atherosclerosis	**0.87**	0.03	0.20	0.10	0.20		**0.80**	-	0.25	0.15	0.28
Acute Myocardial Infarction	**0.83**	0.01	0.04	-	0.30		**0.76**	0.03	0.12	0.03	0.40
Congestive Heart Failure	**0.82**	0.29	0.08	0.07	0.24		**0.79**	0.28	0.21	0.04	0.25
Cardiomyopathy and Structural Heart Disease	**0.77**	0.24	0.10	0.10	0.32		**0.75**	0.24	0.19	0.05	0.34
Heart Valve Disorder	**0.67**	0.19	0.10	0.13	0.49		**0.64**	0.15	0.07	0.19	0.53
Conduction Disorder or Cardiac Dysrhythmia	**0.66**	0.18	0.10	0.14	0.50		**0.62**	0.14	0.09	0.19	0.55
Hypertension	**0.62**	0.17	**0.54**	0.09	0.29		**0.53**	0.08	**0.62**	0.11	0.32
Peripheral Atherosclerosis	**0.57**	0.30	0.17	0.21	0.51		**0.58**	0.14	0.22	0.23	0.54
Cerebrovascular Disease	**0.53**	0.17	0.16	0.25	0.60		**0.54**	0.09	0.14	0.28	0.60
Hyperlipidemia	**0.53**	0.06	**0.69**	0.10	0.23		0.40	-	**0.67**	0.23	0.34
Diabetes Mellitus	0.42	0.23	**0.51**	-	0.51		0.34	0.12	**0.69**	-	0.40
Obesity	0.42	0.15	0.36	0.19	0.63		0.26	0.15	**0.55**	0.09	0.60
Chronic Renal Failure	0.41	**0.60**	0.31	0.03	0.38		0.42	**0.52**	0.39	0.02	0.40
Anemia and other non-Cancer Hematologic Disorders	0.36	**0.73**	0.06	0.16	0.31		0.32	**0.65**	0.1	0.08	0.46
Immunity disorder	0.27	**0.54**	0.04	0.24	0.57		0.28	**0.50**	0.08	0.25	0.60
Other Central and Peripheral Nervous System Disorders	0.26	0.27	0.21	**0.51**	0.55		0.29	0.14	0.19	**0.53**	0.57
Kidney and Vesicoureteral Disorders(excluding renal failure)	0.21	**0.57**	0.38	0.08	0.48		0.17	0.32	0.26	0.26	0.73
Osteoarthritis	0.20	0.11	0.27	**0.61**	0.50		0.23	0.10	0.29	**0.59**	0.51
Chronic Liver Disease(excluding chronic hepatitis)	0.17	**0.53**	0.20	0.16	0.62		0.18	0.43	0.37	0.18	0.61
Other Endocrine	0.17	0.17	**0.50**	0.23	0.64		0.11	0.14	0.34	0.27	0.78
Lupus	0.14	0.33	-	**0.52**	0.60		0.14	**0.48**	-	**0.53**	0.46
Rheumatoid Arthritis	0.14	0.20	-	**0.59**	0.60		0.16	0.33	-	**0.54**	0.58
Chronic Hepatitis	0.12	**0.55**	-	0.13	0.67		0.16	0.47	0.13	0.13	0.72
Back problem	0.12	0.06	0.19	**0.53**	0.66		0.18	0.05	0.14	**0.57**	0.62
Benign Prostatic Hypertrophy(BPH)	0.11	0.17	**0.55**	0.18	0.63		NA	NA	NA	NA	NA
Other Musculoskeletal incl. Osteoporosis	0.10	0.16	0.15	**0.50**	0.69		0.15	0.08	0.15	**0.52**	0.68
Male Genitourinary(GU) exc. BPH	0.07	0.17	**0.52**	0.14	0.67		NA	NA	NA	NA	NA

Female Infertility and GU Disorders	NA	NA	NA	NA	NA		-	0.19	0.08	0.11	0.92

CVMB= Cardiovascular-Metabolic Pattern; HRAI= Hepatorenal-Autoimmune Pattern; MTBU= Metabolic-Urologic Pattern; MTB= Metabolic Pattern; MSK= Musculoskeletal Pattern; CVD= Cardiovascular Diseases Pattern; RAI= Renal-Autoimmune Pattern. NA= Not Applicable to focal gender

Ψ^* = Uniqueness statistic.

§: only positive factor loadings are shown for convenience. We used thresholds of uniqueness<0.70 and factor loading>0.45 to retain chronic conditions.

Table 3 reports the sample prevalence of gender-specific patterns and the prevalence of member chronic conditions within each pattern, stratified by birth-cohorts [see Appendix-D for prevalence of all possible combinations of patterns for each gender with respect to the study sample, and individuals having at least two chronic conditions]. Overall, the most prevalent pattern for men was metabolic-urological (33.9 percent), and the least prevalent was hepatorenal-autoimmune (2.2 percent). For women, the most prevalent pattern was metabolic (24.0 percent), and the least prevalent was renal-autoimmune (1.3 percent). Finally, the most frequently co-occurring two-pattern combination among men was cardiovascular-metabolic and metabolic-urological (20.9 percent), and three-pattern combination was cardiovascular-metabolic, metabolic-urological and musculoskeletal (4.2 percent). For women, the most frequently co-occurring two-pattern combination was metabolic and musculoskeletal (4.0 percent), and three-pattern combination was cardiovascular diseases, metabolic, and musculoskeletal (2.4 percent).

Table 3: Prevalence of Gender-Specific MCC Patterns and their Member Conditions in Commercially Insured Working-Age Adults in Texas, 2008-2013§

MCC Patterns and Member Chronic Conditions	Working-Age Men				Working-Age Women			
	GenX/ Millen.	Baby Boomers		All Birth Cohorts	GenX/ Millen.	Baby Boomers		All Birth Cohorts
		Younger	Older			Younger	Older	
Count of Individuals	95,453	56,275	51,048	202,766	118,157	71,093	60,808	250,058
Pattern–1	_Cardiovascular-Metabolic_(CVMB)				_Cardiovascular Diseases_(CVD)			
Pattern Prevalence, Count(%)	13,471 (14.1)	19,755 (35.1)	25,379 (49.7)	58,605 (28.9)	4,077 (3.5)	7,079 (10.0)	10,455 (17.2)	21,611 (8.6)
Hypertension	92.0	92.9	93.3	92.9	83.1	90.6	93.7	90.7
Hyperlipidemia*	91.2	93.2	94.1	93.1	NA	NA	NA	NA
Coronary Atherosclerosis	11.9	21.0	30.6	23.1	25.4	36.0	41.1	36.5
Acute Myocardial Infarction	3.1	4.7	5.8	4.8	5.5	6.2	6.6	6.3
Congestive Heart Failure	3.1	4.3	6.5	5.0	10.5	10.8	12.4	11.5
Cardiomyopathy and Structural Heart Disease	7.1	8.0	10.4	8.8	17.2	16.6	17.8	17.3
Heart Valve Disorder	8.3	8.5	11.3	9.7	35.2	30.1	31.1	31.6
Conduction Disorder or Cardiac Dysrhythmia	13.9	15.7	20.2	17.2	47.0	39.3	38.5	40.4
Peripheral Atherosclerosis	3.0	5.2	7.8	5.8	9.4	13.2	16.3	14.0
Cerebrovascular Disease	5.2	7.1	10.4	8.1	18.9	21.3	24.6	22.4

Pattern–2	Metabolic-Urologic(MTBU)				Metabolic(MTB)			
Pattern Prevalence, Count(%)	16,612 (17.4)	23,181 (41.2)	28,917 (56.7)	68,710 (33.9)	14,949 (12.7)	21,161 (29.8)	26,157 (43.0)	62,267 (24.9)
Hypertension	79.0	82.9	85.5	83.0	80.7	91.0	93.8	89.7
Hyperlipidemia	87.2	89.2	89.9	89.0	77.0	88.5	92.9	87.6
Diabetes mellitus	32.0	33.8	35.4	34.0	37.3	37.5	38.1	37.7
Obesity*	NA	NA	NA	NA	52.9	34.4	25.9	35.3
Other endocrine*	34.5	31.8	30.5	31.9	NA	NA	NA	NA
Benign Prostatic Hypertrophy(BPH)*	6.2	17.0	28.6	19.3	NA	NA	NA	NA
Male Genitourinary excl. BPH*	10.6	13.4	15.6	13.7	NA	NA	NA	NA
Pattern–3	Musculoskeletal(MSK)				Musculoskeletal(MSK)			
Pattern Prevalence, Count(%)	4,681 (4.9)	5,940 (10.6)	7,957 (15.6)	18,578 (9.2)	8,834 (7.5)	12,195 (17.2)	15,313 (25.2)	36,342 (14.5)
Rheumatoid Arthritis	6.7	7.6	7.7	7.4	15.4	13.7	12.3	13.5
Back problem	64.5	63.8	63.3	63.7	62.8	61.5	57.9	60.3
Osteoarthritis	53.6	67.1	74.2	66.7	51.0	65.4	73.1	65.1
Other Central & Peripheral Nervous System Disorders	55.1	52.9	53.7	53.8	57.9	56.9	51.3	54.8
Lupus	2.0	2.2	1.9	2.0	10.7	8.4	6.8	8.3
Other Musculoskeletal including Osteoporosis	46.1	43.4	43.5	44.1	43.8	48.7	59.5	52.1
Pattern–4	Hepatorenal–Autoimmune(HRAI)				Renal–Autoimmune(RAI)			
Pattern Prevalence, Count(%)	1,084 (1.1)	1,425 (2.5)	1,941 (3.8)	4,450 (2.2)	993 (0.8)	989 (1.4)	1,182 (1.9)	3,164 (1.3)
Chronic Renal Failure	36.9	43.4	47.6	43.7	30.6	37.0	48.4	39.3
Anemia & Other non-Cancer Hematoma Disorders	57.5	58.5	60.1	58.9	77.6	77.1	79.4	78.2
Immunity Disorder	49.2	40.9	37.7	41.5	69.1	68.6	65.0	67.4
Kidney and Vesicoureteral DS (exclude renal failure)*	31.0	32.7	36.2	33.8	NA	NA	NA	NA
Chronic Liver Disease (excl. chronic hepatitis)*	45.2	46.4	42.9	44.6	NA	NA	NA	NA
Chronic Hepatitis*	21.6	24.6	21.3	22.4	NA	NA	NA	NA
Lupus*	NA	NA	NA	NA	41.4	35.8	26.7	34.2

§ Prevalence of gender-specific MCC patterns by Birth Cohorts, in top row, are given with respect to cohort specific total individual count. Prevalence of member conditions are shown with respect to individuals having the focal MCC pattern.

* Prevalence for these conditions are marked as NA(Not Applicable) if they are not part of the respective patterns.

Characterization of Multiple Chronic Condition Patterns

The descriptive statistics for each gender-specific pattern by individual-level and ZCTA-level covariates are shown in Table 4.

Table 4. Summary Statistics of Individual Factors and ZCTA-level Community Factors for Gender-Specific Multiple Chronic Condition (MCC) Patterns Among Commercially Insured Working-Age Adults in Texas, 2008-2013 §										
Variables	**MCC Patterns: Men**				**All Men ‡**	**MCC Patterns: Women**				**All Women ‡**
	CVMB	**HRAI**	**MTBU**	**MSK**		**CVD**	**RAI**	**MTB**	**MSK**	
Number of Individuals	58,605	4,450	68,710	18,578	202,776	21,611	3,164	62,267	36,342	250,058
Individual Level Factors										
Count Based Multimorbidity										
Multimorbidity Count in Anchor Pattern	2.4 (0.7)	2.4 (0.6)	2.7 (1.2)	2.7 (0.9)	--	2.7 (1.1)	2.2 (0.4)	2.5 (0.7)	2.5 (0.8)	--
Multimorbidity Count in All Allocated Patterns	6.4 (3.6)	11.0 (5.2)	6.1 (3.4)	8.0 (4.3)	--	9.7 (4.4)	12.5 (5.5)	7.5 (3.7)	8.6 (4.1)	--
Additional Comorbidity Count (excl. conditions used up in patterns)	3.7 (3.0)	8.6 (5.0)	3.4 (3.1)	5.6 (4.1)	3.0 (3.3)	7.0 (3.9)	10.3 (5.4)	5.0 (3.6)	6.1 (3.8)	3.8 (3.6)
Birth Cohorts										
Gen-X/Millennials	24.2	23.0	25.2	24.4	47.1	24.0	18.9	24.3	31.4	47.3
Younger Baby Boomers	33.7	33.7	32.0	32.0	27.8	34.0	32.8	33.6	31.3	28.4
Older Baby Boomers	42.1	43.3	42.8	43.6	25.2	42.0	48.4	42.1	37.4	24.3
Industry Group										
Construction	3.5	3.5	3.3	3.3	3.2	1.5	1.8	1.7	2.0	1.5
Finance/Insurance/ Real Estate	6.2	6.2	6.5	7.1	10.4	6.8	6.8	6.3	6.1	9.7
Manufacturing	13.6	13.5	12.6	13.1	13.6	8.6	9.0	9.5	9.5	8.8
Mining, Agriculture	5.3	5.4	5.3	4.8	4.8	3.1	3.8	3.6	3.4	2.9
Retail Trade	3.8	3.80	3.6	3.6	4.1	3.0	3.1	3.0	3.0	3.3
Services	33.9	34.1	34.1	33.3	31.5	53.9	50.5	49.0	46.3	49.4
Trans., Comm., Elect/ Gas, & Sanitary Serv.	5.1	5.1	5.3	4.8	4.5	3.2	3.3	3.7	3.7	3.2
Wholesale Trade	11.4	11.6	10.6	11.3	10.1	9.1	8.8	8.5	9.2	7.6
Unknown	17.2	16.8	18.8	18.8	17.8	10.7	13.0	14.7	16.9	13.6
Community-Level Socioeconomic Factors										
Dominant Race/Ethnicity of Area										
Non-Hispanic Whites	70.9	70.6	76.3	67.4	72.8	67.1	68.9	73.8	67.4	73.1
Non-Hispanic Blacks	3.0	3.1	2.7	4.1	3.0	5.2	4.8	3.7	6.2	3.8
Hispanics	26.1	26.4	21.0	28.6	24.2	27.7	26.2	22.5	26.4	23.1
Rurality of Area										
Urban	75.2	74.7	74.9	78.8	77.7	73.0	72.6	75.0	77.7	76.0

Large rural	12.5	12.7	12.2	11.3	11.2	13.8	14.1	12.7	11.9	12.1
Small rural	12.3	12.5	12.9	9.9	11.2	13.2	13.3	12.3	10.4	11.9
Median Household Income of Area										
Quintile-1	12.4	13.1	12.1	10.0	10.8	14.5	14.6	14.8	11.3	11.3
Quintile-2	20.7	20.3	20.6	19.9	19.0	22.8	20.5	22.6	20.5	19.4
Quintile-3	17.5	16.6	17.3	17.1	16.9	18.2	16.3	18.4	17.8	17.1
Quintile-4	23.4	24.9	23.3	23.8	23.7	23.8	23.5	23.7	24	24.4
Quintile-5	26.0	25.1	26.8	29.2	29.6	20.8	25.2	20.6	26.4	27.8
Education- Population Concentration with Bachelor or above Degree										
Lower Quartile	21.5	22.1	19.6	21.9	18.9	26.4	26.6	21.3	23.5	20.0
Middle 50th	51.7	51.9	51.6	52.9	50.9	54.0	52.7	51.9	50.8	51.7
Top Quartile	26.9	26.0	28.9	25.2	30.2	19.6	20.7	26.8	25.7	28.3
Foreign-born Population Concentration										
Lower Quartile	37.4	38.1	39.7	33.6	35.3	39.2	40.3	39.3	34.8	37.3
Middle 50th	48.2	47.5	48.6	49.1	50.1	46.4	45.7	47.9	50.9	49.0
Top Quartile	14.5	14.4	11.7	17.3	14.7	14.4	13.9	12.7	14.3	13.7
Note: The Chi-square were test significant, p-value<0.05 § Summary statistics for all categorical variables are reported as proportions (%), and for continuous variables as mean (standard deviation). ‡ These columns represent distribution of all Men and Women individuals sampled for the study.										

The MCC pattern prevalence for men was nearly double in older boomers compared to Generation-X/Millennials (cardiovascular-metabolic: 42.1 vs. 24.2 percent; hepatorenal-autoimmune: 43.3 vs. 23.0 percent; metabolic-urological: 42.8 vs. 25.2 percent; musculoskeletal: 43.6 vs. 24.4 percent). Whereas, in women, the pattern prevalence growth was dispersed, ranging from 1.2 to 2.6 times (cardiovascular diseases: 42.5 vs. 24 percent; renal-autoimmune: 48.4 vs. 18.9 percent; metabolic: 42.1 vs. 24.3 percent; musculoskeletal: 37.4 vs. 31.4 percent). For all patterns, prevalence by subscribers' employer industry was highest in the service industry (men: 33.3-34.1 percent; women: 46.3-53.9 percent), and lowest in the construction industry (men: 3.3-3.5 percent; women: 1.5-2.0 percent). For each pattern, about one-quarter of individuals lived in the Hispanic dominant areas (men: 21.0-28.6 percent; women: 22.5-27.7 percent), fewer in Black dominant areas (men: 2.7-4.1 percent; women: 3.7-6.2 percent), about three-fourths in urbanized areas (men: 74.7-75.2 percent; women: 72.6-77.7 percent), about one-tenth in small (men: 9.9-12.5 percent; women: 10.4-13.3 percent) and large (men: 11.3-12.7 percent; women: 11.9-14.1 percent) rural areas. For all patterns in both genders, prevalence was higher at bottom quartile and lower at top quartile of area-level foreign-born population concentration.

The associations of each gender-specific MCC pattern with individual-level and area-level factors are reported in Table 5, adjusting for additional comorbidity burden, subscriber type, employer industry, and area-level disability rates. Among men, compared to Generation-X/Millennials, younger boomers have 17-190 percent increased risk of developing cardiovascular-metabolic (adjusted odds ratios [adj. OR]=2.74), hepatorenal-autoimmune (adj.

OR = 1.17), metabolic-urologic (adj. OR = 2.90), and musculoskeletal (adj. OR = 1.60) patterns; and older boomers have 86-372 percent increased risk of developing metabolic-urologic (adj. OR = 4.31), metabolic-urologic (adj. OR = 4.72), and musculoskeletal (adj. OR = 1.86) patterns. Among women, compared to Generation-X/Millennials, younger boomers have 86-250 percent increased risk of developing cardiovascular (adj. OR = 2.14), metabolic (adj. OR = 2.50), and musculoskeletal (adj. OR = 1.86) patterns, but 16 percent reduced risk for renal-autoimmune pattern (adj. OR = 0.84); and older boomers have 153-328 percent increased risk of developing cardiovascular (adj. OR = 3.45), metabolic (adj. OR = 4.28), and musculoskeletal (adj. OR = 2.53) patterns, but 25 percent reduced risk for renal-autoimmune pattern (adj. OR = 0.75).

Table 5. Association of Gender-Specific MCC Patterns with Individual Factors and ZCTA-level Community Factors Among Commercially Insured Working-Age Adults in Texas, 2008-2013. Adjusted Odds Ratios(95% CI) Based on Multilevel Mixed-Effects Logistic Regression §

Variables	**MCC Patterns for Adult Men**				**MCC Patterns for Adult Women**			
	CVMB	**HRAI**	**MTBU**	**MSK**	**CVD**	**RAI**	**MTB**	**MSK**
Individual Factors								
Birth Cohort								
Gen-X/Millennials [*Ref.*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Younger Baby Boomers	2.74***	1.17***	2.90***	1.60***	2.14***	0.84***	2.50***	1.86***
Older Baby Boomers	4.31***	1.02	4.72***	1.86***	3.45***	0.75***	4.28***	2.53***
Community-Level Socioeconomic Factors								
Dominant Race/Ethnicity of Area								
Non-Hispanic Whites [*Ref.*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Non-Hispanic Blacks	1.06	1.39***	1.05	0.89*	1.19***	1.49***	1.43***	0.87***
Hispanics	1.17***	1.07	1.19***	0.81***	1.02	1.13**	1.23***	0.89***
Rurality of Area								
Urban [*Ref.*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Large rural	1.11***	0.89**	1.10***	1.03	1.02	0.96	1.01	1.00
Small rural	1.05*	0.80***	1.03	1.09***	1.01	0.93	0.96	0.99
Education- Population Concentration with Bachelor or above Degree								
Lower Quartile	1.11***	0.97	1.07***	1.04	1.15***	0.97	1.17***	1.02
Middle 50th [*Reference*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Top Quartile	0.89***	0.91*	0.91***	0.96	0.80***	1.1	0.67***	0.99
Community-Level Socioeconomic Factors								
Median Household Income of Area								
Quintile-1 [*Ref.*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Quintile-2	0.93**	1.01	0.97	1.06	0.98	0.92	0.96	1.07**
Quintile-3	0.94**	0.99	0.93**	1.03	0.95	0.85*	0.96	1.05
Quintile-4	0.94*	1.05	0.95*	1.06	0.96	0.93	0.95	1.02
Quintile-5	0.94*	0.95	0.96	1.10**	0.89**	0.95	0.97	1.03
Foreign-born Population Concentration								
Lower Quartile	1.08***	0.88***	1.03*	1.06**	1.04	0.90**	0.99	1.02
Middle 50^{th} [*Ref.*]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Top Quartile	0.95*	1.20***	0.93***	0.92**	1.02	0.91	0.97	0.99
Constant term	0.09***	0.00***	0.13***	0.03***	0.01***	<0.01***	0.09***	0.04***
ZCTA-level Estimated Variance Component of Random Effects	1.03***	1.03**	1.03***	1.02***	1.05***	1.04***	1.05***	1.02***
Wald's Chi^2	25398.5	10075.2	28409.4	14310.6	23882.6	7835.9	29502.1	25995.6

Note: * p<0.10; ** p<0.5; *** p<0.01;
§ Adjusted Odds Ratios with 95% confidence intervals in parentheses are reported from multilevel mixed effects logistic regression models with ZCTA-level robust standard error to account for any misspecification bias. Estimates for control variables are not displayed.

Considering race/ethnicity, men living in Hispanic dominant areas, compared to non-Hispanic White dominant areas, have 17-19 percent increased risk for developing cardiovascular-metabolic (adj. OR = 1.17) and metabolic-urologic (adj. OR = 1.19) patterns but 19 percent reduced risk for musculoskeletal pattern (adj. OR = 0.81). Further, men living in non-Hispanic Black dominant areas have 39 percent increased risk for developing hepatorenal-autoimmune (adj. OR = 1.39) but 11 percent reduced risk for musculoskeletal (adj. OR = 0.89) pattern. Women living in Hispanic dominant areas, compared to non-Hispanic White dominant areas, have 19-49 percent increased risk of developing cardiovascular (adj. OR = 1.19), renal-autoimmune (adj. OR = 1.49), and metabolic (adj. OR = 1.43) patterns, but 13 percent reduced risk for musculoskeletal pattern (adj. OR = 0.87).

Concerning rurality, men living in large rural areas, compared to urban areas, have 10-11 percent increased risk for cardiovascular-metabolic (adj. OR = 1.11) and metabolic-urologic (adj. OR = 1.10) patterns, but 11 percent reduced risk for hepatorenal pattern (adj. OR = 0.89). Further, men living in small rural areas have 5-9 percent increased risk for developing cardiovascular-metabolic (adj. OR = 1.05) and musculoskeletal (adj. OR = 1.09) patterns, but 20 percent reduced risk for hepatorenal-autoimmune pattern (adj. OR = 0.80). For women, there was no statistical evidence of association between MCC patterns and rurality of residential area.

Regarding area-level income, men living in moderate to very high income areas (2nd quintile onwards), compared to those living in lower income areas (1st quintile), have 6-7 percent reduced risk for developing cardiovascular-metabolic pattern (adj. OR = 0.93-0.94), but 10 percent increased risk for musculoskeletal pattern (adj. OR = 1.10) if living in very high income areas (5th quintile). The associations for women were mostly non-significant.

Considering area-level educational attainment, men living in areas with lower college educated population concentration (bottom quartile), compared to areas with moderate college educated population concentration (ZCTAs in the interquartile range), have 7-11 percent increased risk for developing cardiovascular-metabolic (adj. OR = 1.11) and metabolic-urological (adj. OR = 1.07) patterns. Further, men living in areas with higher college educated population concentration (top quartile) have 9-11 percent reduced risk for developing cardiovascular-metabolic (adj. OR = 1.89) and metabolic-urological (adj. OR = 0.91) patterns. Similarly, women living in areas with lower college educated population concentration have 15-17 percent increased risk for developing cardiovascular (adj. OR = 1.15) and metabolic (adj. OR = 1.17) patterns, but those living in areas with higher college educated population concentration have 20-33 percent reduced risk for developing cardiovascular (adj. OR = 0.80) and metabolic (adj. OR = 0.67) patterns.

Finally, considering immigration dynamics, men living in areas with lower foreign-born population concentration, compared to areas with moderate foreign-born population concentration, have 3-8 percent increased risk for developing cardiovascular (adj. OR = 1.08), metabolic-urologic (adj. OR = 1.03) and musculoskeletal (adj. OR = 1.06) patterns, but 12 percent reduced risk for hepatorenal-autoimmune pattern (adj. OR = 0.88). Whereas, men living in areas with higher foreign-born population concentration have 5-8 percent reduced risk of developing cardiovascular (adj. OR = 0.95), metabolic-urologic (adj. OR = 0.93) and musculoskeletal (adj. OR = 0.92) patterns, but 20 percent increased risk for hepatorenal-autoimmune pattern (adj. OR = 1.20). There were no statistically significant associations for women regarding immigrant population concentration.

DISCUSSION AND CONCLUSIONS

This study identifies gender-specific patterns of multiple chronic conditions (MCC) in working-age adults and characterizes their variations with respect to birth-cohorts and community-level socioeconomic determinants to elicit health disparity.

Identification of Multiple Chronic Condition Patterns

Our analyses yielded four gender-specific MCC patterns among working-age adults (men: cardiovascular-metabolic, hepatorenal-autoimmune, metabolic-urologic, and musculoskeletal; women: cardiovascular diseases, renal-autoimmune, metabolic, and musculoskeletal). These patterns closely corresponded with the prior research, although they differ in composition of member conditions: metabolic (Holden et al. 2011; Prados-Torres et al. 2012), cardiovascular (García-Olmos et al. 2012; John et al. 2003), cardiovascular and metabolic pattern (Cornell et al. 2007; Newcomer et al. 2011), musculoskeletal (Rocca et al. 2014; Jackson et al. 2016), and the hepatorenal-immune (Cornell et al. 2007; Wong et al. 2011) patterns.

In our study sample, about 40 percent of men and 34 percent of women had at least one MCC pattern; however, within the subsample of adults with 2+ chronic conditions these prevalence rise to 71 percent and 50 percent respectively. Further, pattern prevalence across birth-cohorts, Generation-X/Millennials to older baby boomers, indicated three-fold growth in men, and two to five-fold growth in women confirming higher onset of chronic conditions with aging. These findings are consistent with prior research (Rocca et al. 2014; Barnett et al. 2012) and highlight the need to customize screening and prevention services based on MCC patterns for working-age adults from different birth-cohorts to help them manage disease burden effectively.

Among working-age men, the prevalence of cardiovascular-metabolic (28.9 percent) and metabolic-urologic (33.9 percent) patterns were consistent with prior research that reported prevalence of 30-39 percent for similar patterns (Aguilar et al. 2015; O'Neill and O'Driscoll 2015; Schäfer et al. 2010). The dominance of hypertension and hyperlipidemia (above 80 and 87 percent respectively) in these patterns across three birth-cohorts suggests two potential pathways of pattern formation for men: developing cardiovascular diseases leading to cardiovascular-metabolic pattern or developing additional metabolic and genitourinary disorders leading to metabolic-urological pattern. Among working-age women, cardiovascular (8.6 percent) and metabolic (24.9 percent) patterns prevalence was lower than their analogous patterns among men. Such gender differences were consistent with prior research reporting lower prevalence for women on similar disease clusters (e.g., O'Neill and O'Driscoll 2015; Schäfer et al. 2010).

Our findings highlight structural differences in the metabolic and renal-autoimmune related patterns between genders. The metabolic disorders (hypertension and hyperlipidemia) clustered with urologic disorders for men only, whereas obesity is integral to metabolic pattern for women only. Next, renal-autoimmune related patterns had three common conditions for both genders- ‘chronic renal failure’, ‘anemia and other non-cancer hematoma disorders’, and ‘immunity disorders’ with the latter two conditions having about 20 percentage points higher prevalence among women for all birth-cohorts. Further, lupus occurred only for women; and three additional conditions ‘chronic hepatitis’, ‘kidney and vesicoureteral disorders’ and ‘chronic liver disease’ occurred only for men.

Finally, the musculoskeletal pattern had similar structure for both gender, but higher prevalence among women for each birth-cohort. Interestingly, consistent with prior research (e.g., Briggs et al. 2016), rheumatoid arthritis had about two-fold higher prevalence among women than men (Generation-X/Millennials: 15.4 vs. 6.7 percent; younger boomers: 13.7 vs. 7.6 percent; older boomers: 12.3 vs. 7.7 percent), and lupus had about 4-5-fold higher prevalence (Generation-X/Millennials: 10.7 vs. 2.0 percent; younger boomers: 8.4 vs. 2.2 percent; older boomers: 6.8 vs. 1.9 percent). However, the remaining four conditions in musculoskeletal pattern were not different.

Association of MCC Patterns with Individual and Community-Level Factors

Our regression analyses provided important insights about variations in the likelihood of developing gender-specific MCC patterns for individuals from different birth-cohorts and geographic areas. Except for renal-autoimmune related patterns, the risks of developing cardiovascular, metabolic, and musculoskeletal patterns among younger and older baby boomers are nearly two to four-fold compared to Generation-X/Millennials, adjusting for other covariates. These findings, consistent with prior research (Rocca et al. 2014; Barnett et al. 2012), suggest that the future chronic burden may rise significantly for Generation-X/Millennials (and the younger boomers) as they age.

Place of residence is an important determinant of population health (Diez Roux 2001; Diez Roux and Mair 2010). Recent research reported association of higher chronic diseases burden with lower socioeconomic status of individuals and their place of residence (Barnett et al. 2012; Diaz et al. 2015; Jackson et al. 2016; Johnson-Lawrence et al. 2017; Lenzi et al. 2016; Lu and Qin 2014; Lynch et al. 2015; Marengoni et al. 2011; Roberts et al. 2015; Violan et al. 2014). Our study extends this body of research by showing the association of MCC patterns among working-age adults with ZCTA-level socioeconomic determinants including minority concentration, rurality, median household income, educational attainment, and foreign-born population concentration. Our results suggest that women living in Hispanic dominant areas are at higher risk of developing renal-autoimmune and metabolic patterns compared to non-Hispanic White dominant areas; and women living in non-Hispanic Black dominant areas are additionally prone to cardiovascular pattern. Likewise, men living in Hispanic dominant areas are more likely to develop cardiovascular-metabolic and metabolic-urologic patterns; and those living in non-Hispanic Black dominant areas are more likely to develop hepatorenal-autoimmune pattern. Coincidentally, both men and women residing in minority dominant areas are less prone to develop musculoskeletal patterns compared to those living in non-Hispanic White dominant areas. These findings could be related to providers' attitude toward pain assessment for their patients often leading to underestimation of musculoskeletal disorders among minorities (Anderson, Green, and Payne 2009; Shavers, Bakos, and Sheppard 2010).

Prior research indicates substantial disparity in multimorbidity for rural residents (e.g., Goeres et al. 2016; Lynch et al. 2015). Our results offer partial support to this strata of disparity literature; and suggests that working-age men living in large or small rural areas, compared to those living in urbanized areas, are at higher risk of developing cardiovascular-metabolic, metabolic-urologic, and musculoskeletal patterns, but less likely to develop hepatorenal-autoimmune pattern. Surprisingly, we did not find any evidence of rural disparity for working-age women.

We found mixed evidence on the relationship between MCC patterns and area-level household median income. Consistent with prior research (e.g. Johnson-Lawrence et al. 2017; Tucker-Seeley et al. 2011), our study indicates lower risk of developing cardiovascular-metabolic and metabolic-urologic patterns among men residing in moderate to high-income areas compared to low-income areas. However, associations for other gender-specific patterns were non-significant.

Educational attainment is regarded as a proxy for an individual's socioeconomic status since it could determine occupational/income opportunities, and (healthier) lifestyle choices (e.g., Johnson-Lawrence et al. 2017; Schiotz et al. 2017; Tucker-Seeley et al. 2011). This stream of research suggests that people with higher education have lower chronic disease burden. Our results extend this literature by considering area-level educational attainment as a potential source of socioeconomic disparity for MCC patterns. Specifically, our results suggest working-age men and women living in areas with higher concentration of college-plus educated population are at lower risk of developing select MCC patterns compared to areas with moderate concentration (men: cardiovascular-metabolic, and metabolic-urologic pattern; women: cardiovascular, and metabolic), and conversely at higher risk if living in areas with lower concentration of college-plus educated populations.

Finally, the state of Texas presents a unique population perspective because it adds about a quarter of million new residents annually through interstate and international migration, most of whom are foreign-born and primarily from Asia and Latin America (White et al. 2015, 2017). Given profound differences in culture, health behaviors and propensities to certain chronic conditions among foreign-born versus domestic population, we suspect "importation" of health gradient may be at play (Buttenheim et al. 2010; Wang and Hu 2013). Our results indicate working-age men living in areas with a higher foreign-born population concentration, compared to areas with moderate concentration, are less likely to have cardiovascular-metabolic, metabolic-urologic, and musculoskeletal patterns but are at higher risk for hepatorenal-autoimmune pattern. The relationship direction was reversed for areas with lower foreign-born concentrations. These results are consistent with prior research comparing MCC prevalence between native populations and foreign-born immigrants (e.g., Diaz et al. 2015; Lenzi et al. 2016; Lu and Qin 2014; Johnson-Lawrence et al. 2017). However, we did not find significant differences for any women-specific patterns.

Limitations

Although our study is one of the first to examine gender-specific MCC patterns for working-age American adults using a long-run claims data, our results need to be interpreted with caution due to several limitations. Our data came from a single insurance carrier from Texas that may not represent the US population or even Texas. Our inclusion criteria eliminated a significant number of enrolled members such as individuals with excessive enrollment gaps and older baby boomers who become Medicare eligible during study period and expected to have more chronic conditions. Finally, our use of ZCTA as unit of abstraction to quantify 'place effect' may be an oversimplification because residents within ZCTA could be heterogeneous on socioeconomic determinants.

Conclusion

Rising prevalence of multiple chronic conditions, with the aging population, poses major challenges to the fragmented US health care system. Despite growing interest in research and practice, our understanding of multiple chronic conditions is still inadequate. Our study sought to address some of the extant knowledge deficits, particularly for the American population, by focusing on working-age adults covered by a commercial carrier and residing in the largest and most diverse states in the contiguous US. Our study offers an initial baseline for gender-specific patterns of multiple chronic conditions and identifies disparity in the risk of developing these patterns with respect to birth-cohorts and area-level socioeconomic determinants.

REFERENCES

NOTE: Only selected references are listed here. Full list is available upon request.

Abad-Díez, J.M., Calderón-Larrañaga, A., Poncel-Falcó, A., Poblador-Plou, B., Calderón-Meza, J.M., Sicras-Mainar, A., Clerencia-Sierra, M., Prados-Torres, A., 2014. Age and gender differences in the prevalence and patterns of multimorbidity in the older population. *BMC Geriatrics*, 14, 75.

Abbass, I., Revere, F.L., Mitchell, J., Appari, A., 2017. Medication nonadherence: Role of cost, community, and individual factors. *Health Services Research*, 52 (4), 1511-1533.

Aguilar, M., Bhuket, T., Torres, S., Liu, B., Wong, R.J., 2015. Prevalence of the metabolic syndrome in the United States, 2003-2012. *JAMA* 313(19), 1973-1974.

Álvarez-Gálvez J, Ortega-Martín E, Carretero-Bravo J, Pérez-Muñoz C, Suárez-Lledó V and Ramos-Fiol B .(2023). Social determinants of multimorbidity patterns: A systematic review. *Frontiers in Public Health*, 11:1081518.

Anderson, K.O., Green, C.R., Payne, R., 2009. Racial and ethnic disparities in pain: causes and consequences of unequal care. *Journal of Pain*, 10(12), 1187-1204.

Badland, H., G. Turrell, and B. Giles-Corti. 2013. "Who Does Well Where? Exploring How Self-Rated Health Differs Across Diverse People and Neighborhoods." *Health & Place*, 22, 82-89.

Barnett, K., Mercer, S.W., Norbury, M., Watt, G., Wyke, S., Guthrie, B. 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *Lancet,* 380 (9836), 37-43.

Basu, J., Avila, R., Ricciardi, R., 2016. Hospital readmission rates in US States: Are readmissions higher where more patients with multiple chronic conditions cluster? *Health Services Research*, 51(3), 1135-1151.

Bierman, A. S., Wang, J., O'Malley, P. G., & Moss, D. K. (2021). Transforming care for people with multiple chronic conditions: Agency for Healthcare Research and Quality's research agenda. *Health Services Research*, 56(Suppl 1), 973.

Bierman, A., Khanna, G. 2019. Learning While Caring: Advancing Patient-Centered Care for People with Multiple Chronic Conditions. Agency for Healthcare Research and Quality, Rockville, MD. Last accessed on 5/15/2022. https://tinyurl.com/y3dodg4w

Briggs, A.M., Cross, M.J., Hoy, D.G., Sànchez-Riera, L., Blyth, F.M., Woolf, A.D., March, L., 2016. Musculoskeletal health conditions represent a global threat to healthy aging: A report for the 2015 World Health Organization World Report on Ageing and Health. *Gerontologist*, 56(Suppl_2), S243-S255.

Buttenheim, A., Goldman, N., Pebley, A.R., Wong, R., Chung, C., 2010. Do Mexican immigrants 'import' social gradients in health to the US?" *Social Science Medicine,* 71, 1268-1276.

Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. Journal of Chronic Diseases 1987;40:373–383.

Cornell, J.E., Pugh, J.A., Williams Jr., J.W., Kazis, L., Lee, A.F.S., Parchman, A.F.S., Zeber, J., Pederson, T., Montgomery, K.A., Noël, P.H., 2007. Multimorbidity clusters: Clustering binary data from multimorbidity clusters: Clustering binary data from a large administrative medical database." *Applied Multivariate Research*, 12 (3), 163-182.

Costello, A.B., Osborne, J.W., 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(7):1-9.

DHHS (U.S. Department of Health and Human Services. 2010. *Multiple Chronic Conditions – A Strategic Framework: Optimum Health and Quality of Life for Individuals with Multiple Chronic Conditions*. Washington, DC.

Diaz, E., Poblador-Pou, B., Gimeno-Feliu, L.A., Calderón-Larrañaga, A., Kumar, B.N., Prados-Torres, A., 2015. Multimorbidity and its patterns according to immigrant origin. A nationwide register-based study in Norway. *PloS ONE*, 10(12), e0145233.

Diez Roux, A.V., 2001. Investigating neighborhood and area effects on health. American *Journal of Public Health*, 91(11), 1783-1789.

Diez Roux, A.V., Mair, C., 2010. Neighborhoods and Health. *Annals NY Academic Science*, 1186(1), 125-145.

Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. Medical Care 1998;36:8–27.

Elixhauser, A., Steiner, C., Palmer, L., 2016. *Clinical Classification Software (CCS) 2015.* U.S. Agency for Healthcare Research and Quality. Washington DC. Last Accessed: 10/05/2019. Available at http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

Ford, E.S., Croft, J.B., Posner, S.F., Goodman, R.A., Giles, W.H., 2013. Co-occurrence of leading lifestyle-related chronic conditions among adults in the United States, 2002-2009. *Preventing Chronic Disease* DOI: dx.doi.org/10.5888/pcd10.120316

Fortin, M., Stewart, M., Poitras, M.E., Almirall, J., Maddocks, H., 2012. A systematic review of prevalence studies on multimorbidity: Toward a more uniform methodology. *Annals of Family Medicine*, 10(2), 142-151.

Freund, T., Kunz, C.U., Ose, D., Szecsenyi, J., Peters-Klimm, F., 2012. Patterns of multimorbidity in primary care patients at high risk of future hospitalization. *Population Health Management*, 15(2), 119-124.

Fujimoto, K., Williams, M.L., Ross, M.W., 2014. A network analysis of relationship dynamics in sexual dyads as correlates of HIV risk misperceptions among high-risk MSM. *Sexually Transmitted Infections*, 91(2), 130-4

García-Olmos, L., Salvador, C.H., Alberquilla, Á., Lora, D., Carmona, M., García-Sagredo, P., Pascual, M., Muñoz, A., Monteagudo, J.L, García-López, F., 2012. Comorbidity patterns in patients with chronic diseases in general practice. *PLoS ONE* 7 (2), e32141.

Glezerman, M., 2016. *Gender Medicine: The Groundbreaking New Science of Gender-and Sex-related Diagnosis and Treatment*. The Overlook Press.

Goeres, L. M., Gille, A.G., Furuno, J.P., Erten-Lyons, D., Hartug, D.M., Calvert, J.F., Ahamed, S. M., Lee, D.S.J., 2016. Rural-urban differences in chronic disease and drug utilization in older Oregonians." *Journal of Rural Health*, 32(3), 269-279.

Hallman, W.K., Kipen, H.M., Diefenbach, M., Boyd, K., Kang, H., Leventhal, H., Wartenberg, D., 2003. Symptom patterns among gulf war registry veterans. Am. J. Pub. Health 93 (4), 624-630.

Herrin, J., St Andre, J., Kenward, K., Joshi, M.S., Audet, A.M.J., Hines, S.C., 2015. Community factors and hospital readmission rates. *Health Services Research*, 50(1), 20–39.

Holden, L., Scuffham, P.A., Hilton, M.F., Muspratt, A., Ng, S.K., Whiteford, H.A., 2011. Patterns of multimorbidity in working Australians. *Population Health Metrics*, 9(1):1.

Huntley, A.L., Johnson, R., Purdy, S., JValderas, .M., Salisbury, C., 2012. Measures of multimorbidity and morbidity burden for use in primary care and community settings: A Systematic Review and Guide. *Annals of Family Medicine*, 10 (2), 134-141.

Jackson, C.A., Dobson, A.J., Tooth, L.R., Mishra, G.D., 2016. Lifestyle and socioeconomic determinants of multimorbidity patterns among mid-aged women: A longitudinal study. *PLoS ONE* 11 (6), e0156804.

John, R., Kerby, D.S, Hennessy, C.H. (2003). Patterns and impact of comorbidity and multimorbidity among community-resident American Indian elders. *Gerontologist*, 43(5), 649.

Johnson-Lawrence, V., Zajacova, A., Sneed, R., 2017. Education, race/ethnicity, and multimorbidity among adults aged 30–64 in the National Health Interview Survey. *SSM-Population Health* 3, 366-372.

Kaiser, H.F. (1974). An index of factor simplicity. *Psychometrika* 39(1), 31-36.

Kirchberger, I., Meisinger, C., Heier, M., Zimmermann, A.K., Thorand, B., Autenrieth, C.S., Peters, A., Ladwib, K.H., Döring, A., 2012. Patterns of multimorbidity in the aged population: Results from the KORA-Age study. *PLoS ONE*, 7 (1), e30556.

Klabunde CN, Legler JM, Warren JL, et al. A refined comorbidity measurement algorithm for claims-based studies of breast, prostate, colorectal, and lung cancer patients. Annals of Epidemiology 2007;17:584–590.

Kline, P. (2014). *An easy guide to factor analysis*. London: Routledge.

Lenzi, J., Avaldi, V.M., Rucci, P., Pieri, G., & Fantini, M.P. (2016). Burden of multimorbidity in relation to age, gender and immigrant status: A cross-sectional study based on administrative data. *BMJ Open*, 6(12), e012812.

Lochner, K.A., Cox, C.S. (2013). Prevalence of multiple chronic conditions among Medicare beneficiaries, United States, 2010. Preventing Chronic Disease 10, E61.

Lu, Y., Qin, L. (2014). Healthy migrant and salmon bias hypotheses: a study of health and internal migration in China. *Social Science Medicine*, 102, 41-48.

Lynch, C.P., Gebregziabher, M., Axon, R.N., Hunt, K.E., Payne, E., Egede, L.E. (2015). Geographic and racial/ethnic variations in patterns of multimorbidity burden in patients with type 2 Diabetes. *Journal of General Internal Medicine*, 30(1), 25-32.

Machlin S.R, Soni, A. (2013). Health care expenditures for adults with multiple treated chronic conditions: Estimates from the Medical Expenditure Panel Survey, 2009. *Preventing Chronic Disease,* DOI: dx.doi.org/10.5888/pcd10.120172

Magnan E. 2015. Algorithm for Identifying Patients with Multiple Chronic Conditions (Multimorbidity). University of Wisconsin - Madison Department of Family Medicine, the University of California - Davis Department of Family and Community Medicine, and the UW Health Innovation Program. Available at: http://www.hipxchange.org/-comorbidities

Marengoni, A., Angleman, S., Melis, R., Mangialasche, F., Karp, A., Garmen, A., Meinow, B., Fratiglioni, L. (2011). Aging with multimorbidity: A systematic review of the literature. *Ageing Research and Review*, 10 (4), 430-439.

Marengoni, A., Bonometti, F., Nobili, A., Tettamanti, M., Salemo, F., Corrao, S., Lorio, A., Marcucci, M., Manucci, P.M. (2010). In-hospital death and adverse clinical events in elderly patients according to disease clustering: The REPOSI study. Rejuvenation Research 13(4), 469-475.

Marengoni, A., Rizzutto, D., Wang, H.X., Winblad, B., Fratiglioni, L. (2009). Patterns of chronic multimorbidity in the elderly population. *Journal of American Geriatrics Society*, 57(2), 225-230.

Naessens, J. M., R. J. Stroebel, D. M. Finnie, N. D. Shah, A. E. Wagie, W. J. Litchy, P. J. Killinger, T. J. O'Byrne, D. L. Wood, and R. E. Nesse. (2011). Effect of multiple chronic conditions among working-age adults. *The American Journal of Managed Care* 17 (2): 118-122.

Newcomer, S.R., Steiner, J.F., Bayliss, E.A. (2011). Identifying subgroups of complex patients with cluster analysis. *American Journal of Managed Care*, 17 (8), e324.

O'Neill, S., O'Driscoll, L. (2015. Metabolic syndrome: a closer look at the growing epidemic and its associated pathologies. *Obesity Review*, 16(1):1-12.

Pett, M. A., N. R. Lackey, and J. J. Sullivan. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage Publications, Inc.

Prados-Torres, A., Poblador-Plou, B., Calderón-Larrañaga, A., Gimeno-Feliu, L.A., González-Rubio, F., Poncel-Falcó, A., Sicras-Mainar, A., Alcalá-Nalvaiz, J.T. (2012). Multimorbidity patterns in primary care, interactions among chronic diseases using factor analysis. *PLoS ONE*, 7 (2), e32190.

Roberts, K.C., Rao, D.P., Bennett, T.L., Loukine, L., Jayaraman, G.C. (2015). Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. *Health Promotion and Chronic Disease Prevention in Canada,* 35 (6), 87-94.

Rocca WA, Boyd CM, Grossardt BR, Bobo WV, Finney Rutten LJ, Roger VL, Ebbert JO, Therneau TM, Yawn BP, St Sauver JL. (2014). Prevalence of multimorbidity in a geographically defined American population: patterns by age, sex, and race/ethnicity. *Mayo Clinic Proceedings*. 89(10):1336-49

Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: Differing perspectives. Journal of Clinical Epidemiology 1993;46:1075–1079

Schäfer, I., Kaduszkiewicz, H., Wagner, H., Schön, G., Scherer, M., van den Bussche, H. (2014). Reducing complexity: A visualization of multimorbidity by combining disease clusters and triads. *BMC Public Health*, 14, 1285.

Schäfer, I., von Leitner, E.C., Schön, G., Koller, D., Hansen, H., Kolonko, T., Kaduszkiewicz, H., Wegscheider, K., Glaeske, G., van den Bussche, H. (2010). Multimorbidity patterns in the elderly: A new approach of disease clustering identifies complex interrelations between chronic conditions. *PloS ONE*, 5 (12), e15941.

Schiotz, M.L., Stockmarr, A., Host, D., Glumer, C., Frolich, A. (2017). Social disparities in the prevalence of multimorbidity – A register-based population study. *BMC Public Health*, 17, 422.

Schultz, S.E., Rothwell, D.M., Chen, Z., Tu, K. (2013). Identifying cases of congestive heart failure from administrative data: A validation study using primary care patient records. *Chronic Disease Injuries in Canada,* 33 (3), 160-6.

Shavers, V.L., Bakos, A., Sheppard, V.B. (2010). Race, ethnicity, and pain among the U.S. adult population. *Journal of Health Care Poor Underserved*, 21(1), 177-220.

Suls J, Bayliss EA, Berry J, Bierman AS, Chrischilles EA, Farhat T, Fortin M, Koroukian SM, Quinones A, Silber JH, Ward BW, Wei M, Young-Hyman D, Klabunde CN. (2021). Measuring Multimorbidity: Selecting the Right Instrument for the Purpose and the Data Source. *Medical Care*, 59(8):743-756.

Sundararajan V, Henderson T, Perry C, et al. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. Journal of Clinical Epidemiology 2004;57:1288–1294

Tucker-Seeley, R.D., Li, Y., Sorenson, G., Subramanian, S.V. (2011). Lifecourse socioeconomic circumstances and multimorbidity among older adults. *BMC Public Health* 11, 313.

van den Akker, M., Buntinx, F., Metsemakers, J.F., Roos, S., Knottnerus, J.A. (1998). Multimorbidity in general practice: Prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases. *Journal of Clinical Epidemiology*, 51 (5) (May), 367-75.

van den Bussche, H., Koller, D., Kolonko, T., Hansen, H., Wegscheider, K., Glaeske, G., von Leitner, E.C., Schäfer, I., Schön, G. (2011). Which chronic diseases and disease combinations are specific to multimorbidity in the elderly? Results of a claims data based cross-sectional study in Germany. *BMC Public Health*, 11, 101.

van Walraven C, Austin PC, Jennings A, et al. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Medical Care 2009;47:626–633.

Violan, C., Foguet-Boreu, Q., Flores-Mateo, G., Salisbury, C., Blom, J., Freitag, M., Glynn, L., Muth, C., Valderas, J.M. (2014). Prevalence, determinants and patterns of multimorbidity in primary care: A systematic review of observational studies. *PloS ONE*, 9 (7), e102149.

Violan, C., Foguet-Boreu, Q., Hermosilla-Pérez, E., Valderas, J.M., Bolíbar, B., Fàbregas-Escurriola, M., Brugulat-Guiteras, P., Muñoz-Pérez, M.Á. (2013). Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. *BMC Public Health*, 13, 251.

Wang, L., Hu, W. (2013). Immigrant health, place effect and regional disparities in Canada. *Social Science Medicine*, 98, 8-17.

Ward, B.W., Schiller, J.S., Goodman, R.A. (2014). Multiple chronic conditions among US adults: A 2012 update. *Preventing Chronic Disease*, 11, E62.

White, S., Potter, L.B., You, H., Valencia, L., Jordan, J.A., Pecotte, B. (2015). *The Foreign-Born Population in Texas: Sources of Growth*. Office of the State Demographer [last accessed on 09/15/2018]. Available at http://demographics.texas.gov/Publications

White, S., Potter, L.B., You, H., Valencia, L., Jordan, J.A., Pecotte, B. (2017). *Texas Migration. Office of the State Demographer* [last accessed on 09/15/2018]. Available at http://demographics.texas.gov/Publications

Wolff, J.L., Starfield, B., & Anderson, G., (2002). Prevalence, expenditures, and complications of multiple chronic conditions in the elderly. *JAMA Archive of Internal Medicine*, 162 (20), 2269-2276.

Wong, A., Boshuizen, H.C., Schellevis, F.G., Kommer, G.J., & Polder, J.J., (2011). Longitudinal administrative data can be used to examine multimorbidity, provided false discoveries are controlled for. *Journal of Clinical Epidemiology*. 64 (10), 1109-1117.

DECISION SCIENCES INSTITUTE

Pedaling towards sustainable commuting

Misa Bakajic
Aalto University
Email: misa.bakajic@aalto.fi

Markku Kuula
Aalto University
Email: markku.kuula@aalto.fi

Niklas Juuranto
Aalto University
Email: niklas.juuranto@aalto.fi

ABSTRACT

This research investigates the effects of a tax-free bicycle fringe benefit on the commuting habits. European decision-makers recognize that cycling can yield numerous benefits such as improved public health, enhanced air quality, and reduced traffic congestion. Finland has implemented its own subsidy scheme to reduce the financial burden associated with bicycle ownership. This study employs a survey methodology, collecting responses from a sample of over 5 000 residents. The findings indicate that the bicycle fringe benefit has positively impacted cycling rates and has the potential to produce intended effects (increased bicycle ownership) and potential unintended effects (transport mode diversification).

KEYWORDS: Cycling, Sustainable transport, Government intervention, Policy

INTRODUCTION

In the summer of 2023, the European Commission will propose an EU Cycling Declaration which aims to serve as a guiding strategic compass for future European rules, funding, and policies (European Commission, 2023). A coalition of 15 EU countries has already endorsed the Declaration urging the recognition of cycling as a complete mode of transportation. Such initiatives aim to make cycling as a more prominent feature of the European transport framework and promise to attract more targeted investment in the future.

The recent push to promote cycling across Europe follows a decades old strategy aiming to take advantage of the benefits of increased bicycle use. Cycling has been linked to improvements in public health, air quality and reduction in traffic congestion in urban environments (Braun, et al 2020; Pucher, 2012). By directing investment into the expansion of cycling infrastructure the use of bicycles become more attractive option of individuals (Buehler & Pucher, 2012). With the arrival of pedelec (pedal electric cycle) and other electric bicycle (e-bike) the travel range of cyclists is increased, allowing for better connection between cities and rural areas (European Commission, 2023). E-bikes also enable cities and regions to better achieve their climate goals (McQueen et al. (2020). Therefore, the rising popularity of the e-bike represents an emerging phenomenon that can alter future mobility strategies in European cities.

In Finland government policy has followed suit with European trends and aimed to stimulate cycling through investments and other interventions (ECMT,2004, p. 44). In 2021 the Finnish government updated the existing bicycle fringe-benefit to make it more attractive and significantly increasing the popularity of the scheme (Hytönen, 2023). As such, the current success of the tax-free bicycle fringe benefit (TFBFB) as a policy intervention tool represents a unique research setting. This study explores the effect of the recently updated bicycle fringe benefit in Finland and evaluate the impact it had on the commuting habits. To achieve the aim empirical data is gathered using the survey method involving 5 269 respondents. In the following section the literature that was used to develop the three hypotheses is first presented. Subsequently, the methodology is explained along with the measures and validity checks performed. That is followed by the findings and related discussion. The conclusion and the limitations of the study are discussed, and future research avenues are proposed as the end.

LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

In recent years policy makers in the Nordics have sought ways to stimulate cycling to take advantage related benefits (Naturvårdsverket, 2019; Ministry of Transport and Communications, 2011). Although regional and local authorities hold the primary responsibility for detailed planning and execution of cycling policies, national-level support is crucial in establishing appropriate legal, regulatory, and financial frameworks. In Finland, the Ministry of Transport and Communications has formulated a Cycling Policy Programme that prioritizes the establishment of a cycling network, with a particular focus on urban areas. The program's objective is to foster cycling and enhance its modal share within the transportation system (Miyake and Crass, 2004, p.28). The strategy aimed at increasing the share of walking and cycling in all the trips made by individuals. In 2011 the initiative set targets for 2020, to increase the share of walking and cycling and to lower trips made by private cars in the same period (Ministry of Transport and Communications, 2011).

To build on the progress made by the Cycling Policy Programme required more specific interventions that targeted bicycle ownership. In 2021, the Finnish government implemented a modification to an existing bicycle fringe to incentivize cycling among the population. The bicycle benefit introduced in Finland was a form of a fringe benefit which would be tax free for up to 1200€ per year (Vero, 2022). Initially, the bicycle benefit was introduced as a model that allowed partial tax deduction. However, it was only in 2021 that it transitioned into a tax-free status (Ministry of Finance, 2021). For individuals, such a change meant that there was a lower cost barrier to bicycle ownership (Bigazzi & Berjisian, 2021).

The decision to introduce the bicycle fringe benefit in Finland, followed other similar government policy inventions across Europe that aimed to influence commuting patterns. Commuting-related fringe benefits come in different forms and can be used to incentivize car ownership, increase the use of public transport, or promote a national bicycle scheme (Nijland & Dijst, 2015). In Europe, bicycles there are over 300 regional and national bicycle scheme which offer incentives ranging from a few hundred euros to a few thousand, depending on the type of bicycle (European Cyclist Federation, 2023). However, despite the popularity of commuting related fringe benefits, they do not always lead to expected outcomes. In some cases, such interventions can have unintended effects that are contrary to the intentions of its creators, making them “perverse incentives” (Tsairi & Katz, 2023). In Finland, however, recent changes to the existing bicycle fringe benefit drastically increased the popularity of the scheme and according to some the number of participants in the scheme is projected to double by the end of 2023 (Hytönen, 2023). However, given that two thirds of the individuals who participated in the current tax-free bicycle fringe benefit were middle-income or higher (Pyöräliito, 2023), policy decision makers have to think about making the scheme more inclusive.

Impact of e-bike adoption

In Sweden, a self-reported study found that when an electrical vehicles subsidy was offered 93% used it to buy electric bicycles. About half of the respondents that acquired e-bikes, did so with the intent to replace their automobiles, while the other half used it to replace other modes of transport (public transport, regular bicycle, walking etc.) (Naturvårdsverket, 2019). Interestingly, that pattern was found to be relevant in both summer and winter months. Like Sweden, the popularity of the bicycle fringe benefit has been significant in Finland and led to a surging popularity of e-bikes (Hytönen, 2023). As e-bikes offer an extended range compared to traditional bicycle models, they can be reliable form of transport even in various kinds of weather. Braun, et al. (2020), found that the number of hot and cold days in a year was not statistically significant as a predictor of bike commuting in large cities. As Finland has comparable weather patterns to Sweden, where the fringe benefit has already resulted in a rise of e-bike popularity, it would be logical to assume that a similar pattern would occur. A survey conducted by the Finnish Cyclist Federation found that the bicycle fringe benefit had increased cycling in both summer and winter months (Pyöräliito, 2023). Hence the first research hypothesis is as follows:

H1= There is an association between fringe benefit bicycle ownership and increased cycling

E-bike effect on automobile usage

The bicycle fringe benefit in Finland had the dual objectives of motivating individuals to improve their health and contribute to the reduction of pollution caused by reducing the use of the automobile for commuting to work (Pyöräliito (2023). Although e-bikes tend to be more expensive than traditional bicycles, they are significantly less costly to own and maintain than automobiles, making them an attractive substitute (Popovich et al. 2014). The transition away from automobiles is not frictionless, however, and must be supported by government investment. As some note, factors such as the availability of adequate biking infrastructure plays a big role in influencing decisions to switch modes of transport (Buehler & Pucher, 2012; Braun, et al. 2020). That issue is being addressed in cities such as Helsinki where significant investments have been made to expand cycling networks connecting the urban core with other areas of the city (ECMT,2004, p. 44). Between 2020 and 2022 alone 41 million euros were allocated to improve cycling infrastructure and expanding (YLE, 2019). Since e-bikes offer high versatility for work commuting and local cycling infrastructure has been improving it would be logical to assume that when cost of e-bike ownership is lowered (through the cycling fringe benefit), there would be a decreased usage of car. Therefore, the second hypothesis is formulated in the following way:

H2= There is an effect between fringe benefit bicycle ownership and decrease of car usage

Evaluating mode choice behavior

Assessing the impact of cycling policies is challenging as it requires data that is often not available (Hong et al. (2020). Having substantial evidence is beneficial in effectively allocating resources and sustaining support for cycling investments, even in cities that have demonstrated a strong dedication to cycling. While it is possible to obtain an aggregate level measurement of cycling traffic, determining the impact of cycling on the actions of individuals is challenging due to transport choice mode behavior. Transport choice mode behavior refers to the decision-making process individuals go through when selecting a mode of transportation for their travel needs. Understanding and predicting transport choice mode behavior from a policy lens can be complicated as it often impacted by both situational and physiological factors (Collins & Chambers, 2005; Rietveld & Daniel, 2004). From a policy perspective, understanding such behavior is an important as it forms a part of the well-established four-step urban travel forecasting model which influenced decision making since before the 1970s (Weiner, 1997, p.43). Understanding how individuals make choices regarding their transportation journey helps to understand and predict traffic in city environments. Although several studies examine the modal shift the automobile to public transport (Collins & Chambers, 2005; Popovich et al. 2014), there has been interest in understanding how increased bicycle usage impacts public transport (Braun, et al. 2020; Hegger, 2007). Considering the assumptions of the previous hypotheses, increased year-round bicycle usage may not only reduce usage of the car but other forms of transport such as buses and metro. As such, the third hypothesis is:

H3= There is an effect between fringe benefit bicycle ownership and decrease of public transport usage

METHODOLOGY

The research aims to gain an in-depth understanding of how the bicycle fringe benefit impact the diversification of transport modes of bicycle owners. As the fringe benefit was only introduced in Finland in recent years, it is a novel phenomenon that has not yet been sufficiently examined. However as outlined in the literature review, key elements of research problem are well understood, such as the tendency of increased bicycle use to impact other modes of transport and the complexity related to understanding transport choice behavior patterns.

The survey method is the most reliable form of data collection method for this research as it allowed for the cross-sectional analysis of a large data set to examine relationships between variables, identify patterns, or compare groups within the surveyed population (Babbie, 2016, p.247). Utilizing the survey method for data collection assumes that answers can used to accurately describe the true characteristics of a respondent sample (avoiding sampling error) and that the chosen participants are an adequate reflection of the target population (avoiding bias) (Fowler, 2013, p. 10). The target population for our study are defined as the people who are already familiar with the tax-free bicycle fringe benefit in Finland. Data was gathered with the assistance of the GoByBike Finland Oy organization, which is private company involved in the bicycle leasing business. This research was not ordered or commissioned by GoByBike, but they were instrumental in dissemination of the survey to existing customers and even offering an incentive for participation. Survey respondents that completed the survey could participate in a draw for a prize worth several hundred euros. As the surveyed group is amongst people who are already more familiar with bicycle benefit and more into cycling than an average person, the survey data is not representative of the entire population.

SendInBlue mailing service (now Brevo) was used to track the survey interaction and responses. The e-mail addresses of the surveyed were provided by GoByBike from their mailing list database. The question types chosen were single answer questions, multiple-choice questions, Likert scale questions and open-ended questions. The survey was open from the 6th of March 2023 up until the 13th of March 2023 and was conducted in Finnish. The survey was made from 5 sections in total, with 16 main questions and two optional feedback questions in the end. Of the 18 135 people the survey was sent to, a total of over 5 269 people responded, so the response rate for the survey was 29.14%. With the surveyed people being from all around Finland. The Chi-Square test for equality of proportions was used to test the hypotheses (Berenson et. al, 2012, p.511).

Ensuring both the reliability and validity of survey data is crucial for producing high-quality and meaningful research results. Reliability refers to the consistency of measurements, while validity refers to the accuracy and meaningfulness of the measurements in relation to the research objectives. To increase reliability the wording in the survey questions was pilot tested several times then reviewed by two of the authors to remove ambiguous language and iron out any issues with the survey design. To increase validity, a brief explanation was added at the start of each section that outlined the focus and explaining the number of related questions.

FINDINGS

The research examines the differences in commuting behavior between individuals who acquired a bicycle using the tax-free cycling fringe benefit and those that did not. While the link between policy intervention and cycling behaviors was present in within the scope of the research scope, causality was not investigated. The findings include in the study come from 5 269 survey respondents. Demographic data was not in the scope of the research and was therefore not collected, however the survey was opened to anyone residing in Finland at the time of the study. The three hypotheses that provide the primary framework for data analysis are structured as follows:

H1= There is an association between fringe benefit bicycle ownership and increased cycling
H2= There is an effect between fringe benefit bicycle ownership and decrease of car usage
H3= There is an effect between fringe benefit bicycle ownership and decrease of public transport usage

Testing of the hypotheses

The chi-square test is used in statistical analysis to determine if there is a significant association or relationship between unpaired categorical variables (Balnaves & Caputi, 2001, p.172-4). It allows researchers to assess whether observed frequencies differ significantly from expected frequencies, helping to identify any discrepancies or patterns in the data. We used the Chi-Square test for equality of proportions (see for example Berenson et. al, 2012, p.511) to test the hypotheses. Based on our survey data, a contingency table was created (Table 1). Our first null hypotheses and the alternative hypotheses are as follows:

H0 = There is no association between fringe benefit bicycle ownership and increase of cycling
H1 = There is an association between fringe benefit bicycle ownership and increase of cycling

Table 1: Contingency table of fridge benefit bicycle ownership effect for increased cycling

Have you increased your cycling during the past three years?	I don't have fringe benefit bicycle	I have fringe benefit bicycle	Sum
No	566 (40.7 %)	825 (59.3 %)	1391 (100 %)
Yes	113 (2.9 %)	3765 (97.1 %)	3878 (100 %)
Sum	679 (43.6 %)	4590 (156.4 %)	5269 (200 %)

Using the values in Table 1, the X^2test value was calculated. Because the X^2 value was 41.9 the null hypothesis is rejected using 0.01 level of significance where the critical value from the chi-square distribution with 1 degree of freedom is 6.64. This confirms that people owning fringe benefit bikes increases cycling. The link between increased cycling and government incentives was found in other studies (Bigazzi & Berjisian, 2021; Naturvårdsverket, 2019). Of the people who do not have a fringe benefit bicycle, only 16.7% have increased cycling. Alternatively, 82% of fringe benefit bicycle owners have increased their cycling.Next the following hypotheses was tested:

H0 = There is no effect between fringe benefit bicycle ownership and decrease of car usage
H2= There is an effect between fringe benefit bicycle ownership and decrease of car usage

Table 2: Contingency table of the fringe benefit bicycle ownership effect for car usage

Have you reduced the use of your car?	I don't have fringe benefit bicycle	I have fringe benefit bicycle	Sum
No	439 (17.4 %)	2083 (82.6 %)	2522 (100 %)
Yes	240 (8.7 %)	2507 (91.3 %)	2747 (100 %)
Sum	679 (26.1 %)	4590 (173.9 %)	5269 (200 %)

As above, using the values in Table 2, the X^2test value was calculated. Because the X^2 value was 3.31 the null hypothesis is accepted using 0.05 level of significance where the critical value from the chi-square distribution with 1 degree of freedom benefit bikes don’t decrease car usage is 3.841. Although there was no statistical confirmed effect between fringe benefit bicycle ownership and decreased car usage. Of the people who do not have a fringe benefit bicycle, 35.3 % have reduced the use of their car while 54.6% of fringe benefit bicycle owners have reduced car usage.As above, the following hypotheses was tested:

H0 = There is no effect between fringe benefit bicycle ownership and decrease of public transport usage
H3= There is an effect between fringe benefit bicycle ownership and decrease of public transport usage

Table 3: Contingency table of fridge benefit bicycle ownership effect on public transport use

Have you reduced the use of public transport?	I don't have fringe benefit bicycle	I have fringe benefit bicycle	Sum
No	490 (16.9 %)	2405 (83.1 %)	2522 (100 %)
Yes	189 (8.0 %)	2185 (92.0 %)	2747 (100 %)
Sum	679 (26.1 %)	4590 (173.9 %)	5269 (200 %)

Similarly, using the values in Table 3, the X^2test value was calculated. Because the X^2 value was 3.69 the null hypothesis is accepted using 0.05 level of significance where the critical value from the chi-square distribution with 1 degree of freedom benefit bikes do not decrease car usage is 3.84. There is no significant statistical effect between fringe benefit bicycle ownership and decreased public transport usage. However, of the people who do not have a fringe benefit bicycle, 27.3 % have reduced the use of public transport, while 47.6 % of fringe benefit bicycle owners have reduced public transport usage.

DISCUSSION

The research investigated the commuting habits of people who did not participate in the tax-free bicycle fringe benefit compared to those that took part. We were able to find that the individuals who took part in the fringe benefit scheme had reported they increase cycling (H1), which was an intended effect of the TFBFB. In addition, although individuals who participated in the bicycle fringe benefit scheme reported that they decreased automobile and public transport usage we were not able to find a significant statistical effect (H2 and H3). Despite that, prior studies have also found that using e-bicycles purchased with the tax-free benefit have decreased automobile usage (H2) (Pyöräliito, 2023). As individuals opt for e-bike instead of automobiles for their daily commute to work, naturally the usage of automobile declines. However, the link between increased cycling and public transport usage (H3) emerges as an unintended effect of TFBFB and its evaluation is less straightforward. As mentioned in the literature review understanding and predicting transport choice mode behavior from a policy lens can be complicated can be complicated as it involves situational and physiological factors. Therefore, assuming that changes in public transit use as the result of increased cycling would be overly simplistic. The link between increased e-bike use and the public transport use within the context of work commute, emerges as salient future research avenue.

Moreover, as the popularity of electric powered bicycles continues to grow, cyclists can now enjoy an extended travel range. The rising popularity of e-bikes in recent years has the potential to elevate cycling as a prominent component within the European transport framework. As government incentive scheme continue to lower the barrier for new e-bike ownership in Finland, they will represent a growing share of traffic. The e-bike is quickly establishing itself as a dependable form of transport for work commuting. Electric bikes often replace the family's third or even second car. Last year, 48,170 electric bikes were sold, while 82,000 new passenger cars were sold, according to the Automotive Information Center (Muoti- ja urheilukauppa, 2023). As the number of e-bikes increases it has the potential to change the fundamental traffic flow characteristics which have been more automobile dominated in recent decades. Another worthwhile future research direction could explore the effects of e-bike popularity on local and national policy decision making. From a government perspective stimulating cycling through policy could be an effective was to encourage healthier lifestyles for citizens. The issue is particularly important in Finland where low physical activity cost the Finnish health system approximately €3.2 billion in 2017 (Kolu et al, 2022). Such investigations would explore the impact of e-bicycles on the future evolution of the Finnish cities, ranging from urban mobility patterns to infrastructure investments.

LIMITATIONS

“Non-additionality" refers to a concept often discussed in the context of policy evaluations, particularly related to interventions or programs designed to achieve specific outcomes. In this context, additionality refers to the extent to which the desired outcomes or impacts of an intervention can be attributed to the intervention itself, beyond what would have occurred in the absence of the intervention. Anderson & Hong, (2022) attempt to evaluate the efficacy of the 2018 Swedish bicycle incentive program by combining administrative, insurance and survey data. However, our study did not aim to demonstrate the efficacy of policy, it relied on self-reported data from people with knowledge of the fringe-benefit who either participated or did not participate in the TFBFB. Even though there was indication that the popularity of the scheme has significantly increased after it was modified (Hytönen, 2023), and even though our findings indicate people who purchased a subsidized bicycle cycled more often, we cannot establish a causal link between the tax-free bicycle benefit and increased cycling. That is mainly because some individuals who purchased e-bikes could have done it even without encouragement of the incentives. Similarly, our data is insufficient to say that people who purchased a bicycle using the tax-free fringe benefit have different transportation habits from those that did not. Even though people with fringe benefits have reduced car usage and people with fringe benefits have reduce public transport usage.

Another limitation of our study is that our target sample consisted of individuals that were source from the Go-Bike mailing list database. As such it cannot be said that the population sample reflects the average population in Finland even though the survey was performed on a national scale. Therefore, our findings include individuals who had prior knowledge of the scheme. Despite this limitation the large survey sample size is provides more accurate and reliable results by minimizing random variability, increasing generalizability, and improving the statistical power of the analysis. The final limitation which should be acknowledged relates to the number of survey questions utilized in the study. These included 16 questions separated in 5 sections, with two additional optional feedback questions that were optional. In addition, the survey did not involve questions relating to demographic data. While the number of questions could have been increased to generate deeper insight, it was decided that a smaller interview length was well suited for a study of this type. The aims were to explore an evolving phenomenon by asking many respondents to get a sense of the trends, rather than investigate the specific actions of individuals or groups.

CONCLUSION

In Europe cycling is emerging as a popular mode of transportation due to its numerous benefits, including improved physical health and reduced traffic congestion. In Finland, the bicycle subsidy scheme has been implemented to promote cycling and incentivize individuals to adopt this sustainable mode of transportation. This research aimed to explore the cycling behavior of individuals who participated in the tax-free bicycle fringe benefit scheme in Finland. By examining the self-reported cycling patterns and habits of respondents who used the scheme versus those that did not, this study offers insight for decision makers who would be involved in crafting future policy. Understanding the link between policy and cycling behavior is of paramount importance, as it can inform policymakers and urban planners in developing strategies to encourage cycling, improve public health, and alleviate traffic congestion. Therefore, the research offers insight that can eventually pave the way for evidence-based interventions and policies promoting cycling as a means of sustainable transportation. The study contributes to the discussion surrounding the factors that impact non-motorized transport modes such as bicycles and e-bikes (Rietveld & Daniel, 2004).

REFERENCES

Anderson, A. & Hong, H., (2022). Welfare Implications of Electric-Bike Subsidies: Evidence from Sweden (No. w29913). National Bureau of Economic Research.

Babbie, Earl R. (2016) *The Practice of Social Research.* 14th ed, Cengage Learning

Balnaves, M. & Caputi, P., (2001). *Introduction to quantitative research methods: An investigative approach*. Sage.

Berenson, M. L., Levine, D. M., & Krehbiel, T. C, (2012), *Basic Business Statistics: Concepts and Applications* 12th Edition, Person Harlow, 889p.

Bigazzi, A. & Berjisian, E., (2021). Modeling the impacts of electric bicycle purchase incentive program designs. *Transportation Planning and Technology*, 44(7), pp.679-694.

Buehler, R. & Pucher, J., (2012). Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*, *39*, pp.409-432.

Braun, L.M., Rodriguez, D.A., Cole-Hunter, T., Ambros, A., Donaire-Gonzalez, D., Jerrett, M., Mendez, M.A., Nieuwenhuijsen, M.J. and de Nazelle, A., (2016). Short-term planning and policy interventions to promote cycling in urban centers: Findings from a commute mode choice analysis in Barcelona. *Transportation Research Part A: Policy and Practice*, 89, pp.164-183**.**

Collins, C.M. & Chambers, S.M., (2005). Psychological and situational influences on commuter-transport-mode choice. *Environment and behavior*, *37*(5), pp.640-661.

ECMT (2004*), Implementing Sustainable Urban Travel Policies: Moving Ahead: National Policies to Promote Cycling*, OECD Publishing, Paris, Available at https://doi.org/10.1787 /9789282103296-en. [Accessed: January 25, 2023].

European Cyclist Federation (2023) Money for bikes: Tax incentives and purchase premiums for cycling in Europe [online]. Available at: https://ecf.com/resources/financial-incentives [Accessed: May 25, 2023].

European Commission (2023). Executive Vice-President Frans Timmermans keynote speech at the Cycling Industries Europe Summit 2023 [online]. Available at: https://ec.europa.eu /commission/presscorner/detail/en/SPEECH_23_1561 Accessed: May 25, 2023].

Fowler Jr, F.J., (2013). *Survey research methods*. Sage publications.

GoByBike (2023) Teemme pyöräedusta kaikille helppoa ja yksinkertaista [We make bicycle benefits easy and simple for everyone] [online]. Available at: https://www.gobybike.fi/ Accessed: May 25, 2023].

Hegger, R., (2007). Public transport and cycling: living apart or together? *Public Transport International*, *56*(2).

Hong, J., McArthur, D.P. & Livingston, M., (2020). The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation*, *47*, pp.2859-2872.

Hytönen, T. (2023) 1 200 euron veroetu ajaa suomalaisia hankkimaan työsuhdepyöriä – löysimme yrityksen, jossa suosio räjähti käsiin [A tax benefit of 1,200 euros drives Finns to buy work bikes - we found a company where the popularity exploded] YLE News [online]. Available at: https://yle.fi/a/74-20030641 [Accessed: January 25, 2023].

Kolu, P., Kari, J.T., Raitanen, J., Sievänen, H., Tokola, K., Havas, E., Pehkonen, J., Tammelin, T.H., Pahkala, K., Hutri-Kähönen, N. & Raitakari, O.T., (2022). Economic burden of low physical activity and high sedentary behaviour in Finland. *Journal of Epidemiology and Community Health*, 76(7), pp.677-684.

McQueen, M., MacArthur, J. & Cherry, C., (2020). The E-Bike Potential: Estimating regional e-bike impacts on greenhouse gas emissions. *Transportation research part D: transport and environment*, 87, p.102482.

Ministry of Transport & Communications [Liikenne- ja viestintäministeriössä] (2011). Strategy for walking and cycling released online]. Available at: https://lvm.fi/en/-/strategy-for-walking-and-cycling-released-785763 [Accessed: January 25, 2023].

Ministry of Finance [Valtiovarainministeriö]. (2021) Lausunto hallituksen esityksestä eduskunnalle laeiksi vuoden 2021 tuloveroasteikosta sekä tuloverolain muuttamisesta ja väliaikaisesta muuttamisesta sekä tulotietojärjestelmästä annetun lain 13 §:n muuttamisesta, Eduskunta- Parliament of Finland. Available at: https://vm.fi/paatos?decisionId=0900908f806dff6a [Accessed: January 25, 2023].

Miyake, M. & Crass, M., (2004) Implementing Sustainable Urban Travel Policies: Moving Ahead: National Policies to Promote Cycling [online]. European Conference of Ministers of Transport.

Muoti- ja urheilukauppa (2023) Jo 40.000 suomalaista polkee työsuhdepyörällä [online] Available at:https://muotijaurheilukauppa.fi/2023/01/12/jo-40-000-suomalaista-polkee-tyosuhdepyoralla/ [Accessed: January 25, 2023].

Naturvårdsverket [Environmental Protection Agency] (2019) Elcykling – vem, hur och varför? [online report]. Available at: https://www.naturvardsverket.se/globalassets/media/publikationer-pdf/6800/978-91-620-6894-3.pdf [Accessed: January 25, 2023].

Nijland, L. & Dijst, M., (2015). Commuting-related fringe benefits in the Netherlands: Interrelationships and company, employee and location characteristics. *Transportation Research Part A: Policy and Practice*, 77, pp.358-371.

Popovich, N., Gordon, E., Shao, Z., Xing, Y., Wang, Y. & Handy, S., (2014). Experiences of electric bicycle users in the Sacramento, California area. *Travel Behaviour and Society*, 1(2), pp.37-44.

Pyöräliito (2023). Polkupyöräetukysely työntekijöille 2023 [Bicycle benefits survey for employees 2023] [online] Available at: https://pyoraliitto.fi/wp-content/uploads/polkupyoraetukysely-tyontekijoille-2023-raportti.pdf [Accessed: January 25, 2023].
Rietveld, P. & Daniel, V., (2004). Determinants of bicycle use: do municipal policies matter? *Transportation research part A: policy and practice*, 38(7), pp.531-550.

Tsairi, Y., Tal, A. & Katz, D., (2023). Policy dissonance when wearing the “employer hat”: The practice of governmental vehicle reimbursement among public sector employees. *Case Studies on Transport Policy*, *12*, p.100997.

Vero. (2022) Fringe benefits from employment [online]. Available at: https://www.vero.fi /en/individuals/tax-cards-and-tax-returns/income/earned-income/fringe-benefits-from-employment/ [Accessed: January 25, 2023].

Weiner, E. (1997) *Urban transportation planning in the United States: An historical overview*. US Department of Transportation.

YLE. (2019) Finland's green scheme to invest €40m in cycling and walking. [online] Available at: https://yle.fi/a/3-10905540 [Accessed: January 25, 2023].

DECISION SCIENCES INSTITUTE

Performance Evaluation of United States Nursing Homes During the COVID-19 Pandemic

Mehrdad Jalali Sepehr
The University of Toledo
Email: Mehrdad.jalalisepehr@rockets.utoledo.edu

Yue Zhang*
The University of Toledo
Yue.Zhang@utoledo.edu

ABSTRACT

The primary purpose of this study is to analyze and group 14073 Nursing Homes (NHs) in the United States during COVID-19. Data Envelopment Analysis (DEA) and subsequently two-step cluster analysis have been implemented as approaches for examining the performance of the NHs. The results show that 74 NHs out of 14073 are efficient. Also, via two-step cluster analysis, three clusters have been identified as the optimum number of clusters for the 14073 efficiency values. Qualitative labels have been assigned to the clusters based on their performance, namely majority, medium-performance, and high-performance.

KEYWORDS: Healthcare analytics, Nursing homes, Data Envelopment Analysis, Clustering, Performance measurement.

INTRODUCTION

Healthcare is one of the largest industries in the world and faces numerous managerial (Hill & Powell, 2009; Nicola et al., 2020). There is an urgent need to apply rigorous tools from other disciplines to healthcare specific environments, and ultimately to improve the quality of healthcare services provided. Nursing homes (NHs), like the rest of the healthcare business in the United States (U.S.), are under growing pressure to enhance efficiency. NHs are a major, expensive, and expanding segment of the entire healthcare market. Nurses play a critical role in healthcare delivery, and nursing leadership is still one of the most significant themes in its infancy (de Vries & Curtis, 2019). Nurses advocate for health promotion, educate patients and the public on the prevention of illness and injury, provide care during acute care episodes, participate in rehabilitation, and provide patient and family support (Abrams et al., 2020). Also, NHs have high contagious disease transmission rates due to overpopulation, common public restrooms, and socializing in common areas, as well as poor infection prevention and control (Davidson & Szanton, 2020).

Multiple sectors, including healthcare, which has been on the front lines of the COVID-19 outbreak, have been radically altered by the pandemic catastrophe (De Luca & Calabrò, 2020; Kim et al., 2022). Therefore, NHs have been one of the most vulnerable sectors in the healthcare system, which faced numerous obstacles during the pandemic, including a shortage of personnel and equipment, medications, and severe psychological and behavioral issues (Mangialavori et al., 2022; Watts et al., 2021). On the other hand, the application of analytics and decision making in the healthcare sector would create substantial values for its stakeholders (Singh et al., 2021); and this issue has become increasingly prevalent because of COVID-19 and increasing ambiguity and serious damages (Gharoie Ahangar et al., 2020; Soto-Ferrari et al., 2022).

In this regard, a comprehensive evaluation of U.S. nursing facilities has been conducted in this study. The efficiency of 14073 U.S. nursing homes have been evaluated by using Data Envelopment Analysis (DEA) and subsequently two-step cluster analysis. It is noted that this efficiency evaluation is based on the specified indicators (Section 3.1) and indicates how efficiently U.S. nursing homes could use their resources as inputs, such as vaccines, beds, employees, etc., to reduce the number of undesirable outputs, such as COIVD-19 cases, and its associated mortalities. Therefore, it is an estimation for their performance during COVID-19 and is not a general performance evaluation for U.S. nursing homes.

This study advocated using healthcare analytics to gain knowledge about the functioning of U.S. nursing homes during a pandemic. We used the Nursing Home Care Compare data from the Centers for Medicare & Medicaid Services (CMS) for this study. This dataset is associated with each registered nursing facility; however, we reported the results at the state level following our analysis. The attained results and the approach in which they were presented would assist in shaping policy and decision making in this sector. For instance, by using the percentage of efficient NHs in each state, policy makers would gain an insight about the performance of NHs in each state, and by comparing them they may prioritize their budget assignment and improvements plans. In addition, following clustering, we categorized the clusters based on their performance: majority, medium performance, and high performance. As a result, it reveals how many U.S. nursing facilities had high efficiency during COVID-19 based on CMS-determined standard indicators.

The research presents an approach of using healthcare analytics for knowledge discovery in the NH sector with the purpose of providing insights for better decision making.

In the following, Section 2 presents a literature review about the performance measurement of NHs. In Section 3, we illustrated the methodology of this study as well as the dataset and indicators. Section 4 presents results, explanations, and interpretations. Finally, Section 5 concludes the paper with discussions about implications of the research and lessons learned during the pandemic.

LITERATURE REVIEW

Kleinsorge & Karney (1992) was one of the first pioneering articles about the use of DEA to measure the efficiency of NHs. Motivated by pressures to improve both efficiency and quality of care in the NH industry, their research studied how the efficiency ratings of a set of NHs are affected when quality measurements are incorporated in overall evaluations, and they employed DEA to assist in this regard, and 22 Kansas NHs were evaluated by DEA. In the study, they first evaluated the NHs in terms of financial conditions, which means they first evaluated the NHs by considering the previously described financial and economic inputs and outputs, and then they added quality metrics to their analysis. Using DEA, Björkgren et al. (2001) evaluated the efficiency of 64 long-term NHs in Finland. In addition to NHs, measuring the performance of visiting nursing services by DEA is an auxiliary and common service for NHs that have been conducted by Kuwahara et al. (2013).

Chattopadhyay & Heffley (1994) evaluated 140 Connecticut NHs by using DEA. Dervaux et al. (2006) used DEA to address two policy challenges in the French nursing care industry. The first challenge was about the NH size and returns to scale, while the second was about the potential impacts of switching from a flat tariff to one based on the severity of case-mix. Iparraguirre & Ma (2015) used three-stage DEA, Stochastic Frontier Analysis (SFA), and a benchmarking approach to investigate efficiency in the provision of social care for older people in 148 English local councils in 2009 and 2010, in terms of the achieving of social-care related quality of life as reported by older users of their social care services.

According to the past research, operating costs are often used as inputs, while output measurements often represent both quantity and quality of services delivered. In this sense, Shimshak et al. (2009) used DEA to assess the three techniques to incorporate quality of care in the evaluation of 38 nursing homes located in central Massachusetts. They considered the separation and combination of the quality and operational efficiencies in their study. Despite significant pension service expansion, China's low care resource utilization persists. Few national and longitudinal studies have examined care service consumption with nursing home production, Zhang et al. (2021) used China Civil Affairs statistics yearbook data to examine NH efficiency and productivity (2012-2016). DEA and the Tobit model were used for their analysis. Min et al. (2016) assessed 2267 U.S. nursing homes in 2010. Using DEA and multilevel modeling, their goal was to find organizational elements that improve NH efficiency. Garavaglia et al. (2011) investigated the efficiency and quality of care in 40 NHs in Italy using two-stage bootstrapped DEA.

Chu et al. (2021) used DEA and Malmquist index (MPI) to model the number of nursing home beds, fixed assets, and medical personnel as input variables and the number of self-care, partial self-care, bed-ridden, and nursing home income as output variables to assess nursing home efficiency. Tran et al. (2019) reviewed nursing home efficiency literature. It demonstrates that DEA and SFA are the most used efficiency measurement methodologies in this field. Ni Luasa et al. (2018) used input-oriented two-stage bootstrap DEA approach to investigate technical and scale efficiency in 39 state and 73 private Irish nursing institutions. In 2018 Dual evaluated technical efficiency of 338 nursing homes in California from 2009 through 2013 using uses two-stage DEA analysis.

According to previously published content, DEA has implemented public policy at both the national and international levels (Sinimole, 2012); and to the best of the authors' knowledge, there has been no research particularly addressing the performance of nursing facilities during COVID-19 by evaluating measures that are specifically focused on pandemic crisis management in nursing homes. Furthermore, while the hybrid DEA-clustering methodology has been employed in other domains (Jalali Sepehr et al., 2019), (Li et al., 2022); it has not been applied to the nursing home industry. Furthermore, the national scale and knowledge discovery approach of this research, which evaluates nursing homes using detailed data from each individual DMU (micro-level) in the U.S. and then presents the results at the state level, would be extremely useful for other researchers and macro-level decision makers in this sector.

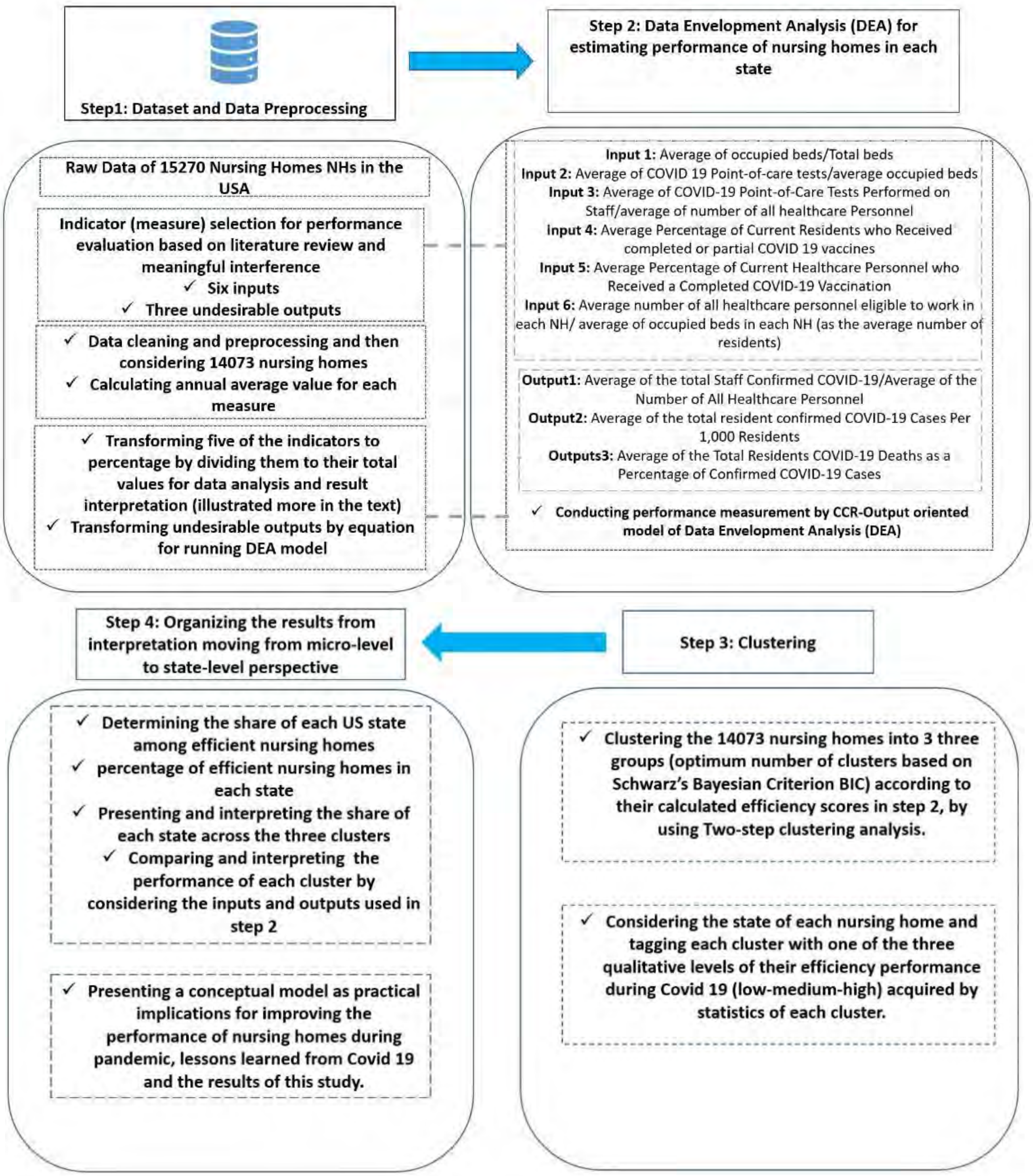
METHODS

This study aims to examine the performance of NHs in the U.S. during COVID-19. In this research, an analytics approach has been adopted to analyze the CMS data by using DEA and two-step cluster analysis. By considering the states of the NHs, our results have been presented at the national scope and state-by-state level.
Initially, using DEA, the efficiency of NHs has been analyzed based on six inputs and three outputs. Second, the optimal number of clusters for calculated efficiency scores has been identified by using two-step cluster analysis. Concerning that, NH efficiencies have been grouped into three clusters. Then, the share of each state in the cluster of efficient NHs (i.e., those with an efficiency score of one), has been identified after clustering. Next, for each cluster, a qualitative label has been assigned, *majority, medium-performance, or high-performance*, and the share of each state in each of these clusters has been reported. As a result, the states with the outstanding performance have a larger share in the high-performance cluster. Figure 1 displays the four steps in this research.
By reviewing the datasets which have been released on the website of the CMS, it has been found that there would be three datasets for each year in which NHs have been struggling with COVID-19 (2020, 2021, and 2022). By reviewing the data, we have considered the dataset of the 2021 for this analysis. The main reason is that 2021 is the year in which a systematic vaccination was started, and we suppose there would be more accurate data about COVID vaccinations for this dataset. Also, the data for 2022 has been in progress yet, and thus it has been considered that it would be better to use a completed dataset. thus there would be about 52 observations per NH in 2021. However, there are some missing data; thus, after indicator selection and removing the weeks for which we had missing values, 14073 NHs in the U.S. have been remained for our analysis.
The indicator selection for this study is based on the previous published content, and their meaningfulness for an analytics approach which aims to discover knowledge of NH performances during the pandemic crisis. In this regard, we have chosen:

1. Total number of all occupied Beds,
2. Number of all beds,
3. COVID-19 Point-of-Care Tests Performed on Residents Since Last Report,
4. COVID-19 Point-of-Care Tests Performed on Staff and/or Personnel Since Last Report,
5. Total Staff Confirmed COVID-19,
6. Total Resident Confirmed COVID-19 Cases Per 1,000 Residents,
7. Total Residents COVID-19 Deaths as a Percentage of Confirmed COVID-19 Cases,
8. Percentage of Current Residents who Received a Completed or Partial COVID-19 Vaccination at Any Time,
9. Number of All Healthcare Personnel Eligible to Work in this Facility for At Least 1 Day,

10. Percentage of Current Healthcare Personnel Who Received a Completed COVID-19 Vaccination at Any Time.

Figure 1: Framework of the methodology in this research



For the dataset of 2021, there are 794020 observations (rows), which are approximatelyrelated to 15270 NHs in the U.S. The data of each NH had been recorded weekly, and
As mentioned earlier, the data are recorded weekly. After removing the weeks with missing data, the average value over the year for each of the 10 indicators for each NH has been considered. Further, to consider the percentage as the scale of the indicator, we calculated some ratios (fractions) and then grouped them into inputs and outputs for the DEA analysis as follows.
The input indicators in our study include:

1. Average of total number of all occupied Beds / average of number of all beds,
2. Average of COVID-19 Point-of-Care Tests Performed on Residents Since Last Report/ average of total number of all occupied Beds,
3. Average of COVID-19 Point-of-Care Tests Performed on Staff and or Personnel Since Last Report / average of number of all healthcare Personnel Eligible to Work in this Facility for At Least 1 Day,
4. Average Percentage of Current Residents who Received a Completed or Partial COVID-19 Vaccination at Any Time,
5. Average Percentage of Current Healthcare Personnel who Received a Completed COVID-19 Vaccination at Any Time,
6. Average number of all healthcare personnel eligible to work in each nursing home for at least one day of this week / average of occupied beds in each nursing home (as the average number of residents).

The output indicators include:

1. Average of the total Staff Confirmed COVID-19 / Average of the Number of All Healthcare Personnel Eligible to Work in this Facility for At Least 1 Day,
2. Average of the total resident confirmed COVID-19 Cases Per 1,000 Residents as percentage,
3. Average of the Total Residents COVID-19 Deaths as a Percentage of Confirmed COVID-19 Cases.

It is noted that DEA would consider inputs and outputs like a production system with a desirable aim to increase outputs. However, all the three outputs identified above are undesirable outputs, because NHs do not want an increase in COVID-19 cases and its associated death. Hence, for DEA, the values of these outputs must be transformed by Equation 1 presented in the next section.

Data Envelopment Analysis

To evaluate the performance of NHs, DEA has been used as an approach for performance measurement (Cooper et al., 2011). DEA has evolved over the last 40 years into a sophisticated quantitative, analytical tool for measuring and evaluating performance. DEA has been properly applied to a broad spectrum of entities engaged in a variety. DEA is a "data-oriented" method of assessing the performance of a group of peer entities known as Decision-Making Units (DMUs) which transform multiple inputs into multiple outputs. In this paper, U.S. nursing homes are the DMUs, and the inputs and outputs have been explained in Section 3.1.

Suppose there are *n* DMUs to be evaluated. Each DMU takes variable amounts of *m* distinct inputs to generate *s* distinct outputs. DMU *j* consumes x_{ij} of input *i* and creates y_{rj} of output *r*. We assume that $x_{ij} \geq 0$ and $y_{rj} \geq 0$ and further assume that each DMU has at least one positive input and one positive output value. Efficiency for DMU *j*, θ_j, is defined as the ratio of a weighted sum of outputs to a weighted sum of inputs as below,

$$\theta_j = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}, \quad (1)$$

where u_r and v_i are non-negative values, denoting the weights associated with output *r* and input *i*, respectively. To make it comparable for different DMUs, the efficiency is defined on a scale of 0-1, i.e., $0 \leq \theta_j \leq 1$.

Unlike statistical or parametric methods, the weights to both outputs and inputs for each DMU are determined by DEA so as to maximize its own efficiency. Therefore, DEA can be viewed as a “self-appraisal” mode, rather than a “peer-appraisal” mode. There are a variety of DEA specifications to determine the weights to outputs and inputs, including input oriented or output oriented models. In particular, the multiplier form of the output oriented CCR model (Mardani et al., 2017), for maximizing the efficiency of DMU *j*, is presented as follows:

$$\min_{u_r, v_i} \sum_{i=1}^{m} v_i \, x_{ij}$$

Subject to:

$$\begin{aligned} &\sum_{r=1}^{s} u_r \, y_{rk} - \sum_{i=1}^{m} v_i \, x_{ik} \leq 0 \qquad \forall\, k = 1, \cdots, n \\ &\sum_{r=1}^{s} u_r y_{rj} = 1 \\ &u_r, v_i \geq 0 \end{aligned} \quad (2)$$

By solving the above model, the weights to outputs and inputs for maximizing the efficiency of DMU *j* as well as the efficiency score can be determined. In order to find the weights for all the DMUs, the above model needs to be solved for *n* times.

The output oriented CCR model would calculate the efficiency for each DMU; however, in this study, the outputs are undesirable, therefore, to transform the outputs to virtual desirable outputs, Equation (3) would be used. By this method, all the three outputs would be multiplied by (-1) and then added to t_r, that is the maximum real value of each undesirable output plus 1. By doing this, the values are transformed. Thus, the DMU with the highest undesirable output would be transformed to the one with the lowest value, and vice versa (Seiford & Zhu, 2002).

$$y_{rj}^{-b} = -y_{rj}^{b} + t_r \geq 0 \quad (3)$$

Two-Step Cluster Analysis

Two-step cluster analysis is an exploratory approach for revealing clusters inside a dataset that might otherwise go unnoticed. This method has numerous beneficial characteristics that set it apart from typical clustering techniques. Three main advantages of two-step cluster analysis are handling of categorical and continuous variables, automatic selection of number of clusters, and scalability.

In this study, to find subgroups of the NHs based on the clustering variable of the efficiency scores obtained from the DEA analysis described in Section 3.2, a two-stage cluster analysis was accomplished with IBM SPSS 27. In this sense, the methodology provides a hybrid method that integrates the findings of the DEA analysis with clustering as a means of carrying out the process of knowledge discovery.

The cluster analysis that involved two steps used two different approaches. The initial phase, known as pre-clustering, involved the separation of groups by the construction of a clustering feature tree that included a distance measure based on the Euclidean distance and the log-likelihood technique(Kent et al., 2014). Second, the ideal subgroup model was decided upon by employing a probabilistic strategy, which is conceptually analogous to latent class analysis, and making use of Schwarz's Bayesian Criteria (BIC) as the clustering criterion (Rundle-Thiele et al., 2015; Tkaczynski, 2016).

RESULTS

Table 1 shows the descriptive statistics for our data, including the minimum and maximum values as well as the mean and standard deviation for each indicator.

Table 1: Descriptive Statistics for inputs, outputs, and efficiency scores

Descriptive Statistics

	N	Minimum (U.S. state or territory)	Maximum (U.S. state or territory)	Mean	Std. Deviation
COVID-19 point of care/average occupied beds (residents)	53	0 (DC)	1.27 (GU)	0.3536	0.18676
occupied beds/total bed	53	0.54 (PR)	0.95 (GU)	0.733	0.07559
COVID-19 test/average of all staff	53	0 (DC)	1.47 (GU)	0.4728	0.25073
Percentage of Current Residents who Received a Completed or Partial COVID-19 Vaccination at Any Time	53	73.44 (NV)	95.89 (VT)	87.8887	4.60284

Percentage of Current Healthcare Personnel Who Received a Completed COVID-19 Vaccination at Any Time	53	53 (MO)	91.76 (HI)	69.7453	10.81813
Number of personnel/ occupied beds	53	1.35 (GA)	3.79 (AK)	1.897	0.40249
Total Resident Confirmed COVID-19 per 1000 residents %	53	0.04 (GU)	0.84 (OK)	0.5887	0.15917
Average of staff COVID-19 confirmed/total staff	53	0.07(HI)	0.56 (AR)	0.3585	0.09785
Total Residents COVID-19 Deaths as a Percentage of Confirmed COVID-19 Cases	53	0 (GU)	39.85 (RI)	21.1142	8.02889
Mean of efficiencies	53	0.4 (GU)	0.72 (LA)	0.5936	0.05456

The efficiency scores of all the 14073 NHs have been calculated after running the output oriented CCR model of DEA. Due to the large number of DMUs, the efficiency score of each individual DMU has not been presented here, and the analysis would be only focused on the performance of all the NHs in each state during the COVID-19 in 2021.

Table 2 displays the average of the estimated efficiency scores for all the NHs in each state. Louisiana (LA) with 0.721 is on the top followed by Puerto Rico (PR) and Missouri (MO).

Table 2: Average of efficiencies for NHs in each state and territories

States/Territories	Average of Efficiencies	States/Territories	Average of Efficiencies
LA	0.721	PA	0.585
PR	0.698	MD	0.584
MO	0.697	MS	0.584
TX	0.688	MI	0.582
OK	0.683	CO	0.581
IL	0.674	UT	0.580
MT	0.650	MA	0.575
DC	0.647	KS	0.572
AR	0.640	VA	0.571
WY	0.639	NE	0.569
NM	0.634	OR	0.568
CA	0.631	IA	0.563
TN	0.628	WA	0.560
GA	0.627	NH	0.558
IN	0.618	SD	0.553

NC	0.615	NY	0.553
AZ	0.606	DE	0.553
SC	0.606	VT	0.552
OH	0.604	AK	0.544
RI	0.602	MN	0.541
NV	0.600	WV	0.538
FL	0.596	ME	0.535
WI	0.596	CT	0.529
KY	0.590	ND	0.527
AL	0.590	HI	0.526
ID	0.590	GU	0.402
NJ	0.590		

Table 3 shows the share of each state among efficient NHs and percentage of efficient NHs in each state. The first column of Table 3 presents the abbreviation of each state that has had efficient NHs according to the purpose of this research (only during COVID-19 in 2021), the selected metrics, and the used analytics approach.
Table 3 reveals that 19 out of 53 states in the U.S. have efficient NHs. The share of each state among efficient NHs has been calculated by dividing the number of efficient NHs in each state by the total number of efficient NHs in the U.S., which are 74 NHs out of 14073. In this regard, Missouri (MO) with 15 efficient NHs out of 74 (20%), is the state with the greatest number of efficient NHs, followed by Texas (15%) and Oklahoma (9%).
Column 3 shows the percentage of efficient NHs in each state, which was calculated by dividing the number of efficient NHs in each state by the total number of NHs in that state. In this regard, Nevada (NV) has the greatest value with 3 efficient NHs out of 64, followed by Oklahoma with 7 efficient NHs out of 276.

Table 3: Share of each state among efficient nursing homes and percentage of efficient nursing homes in each state

States	The share of each state among efficient nursing homes	Percentage of efficient nursing homes in each state
CA	5%	1%
CO	1%	0%
FL	5%	1%
GA	1%	0%
IL	8%	1%
IN	4%	1%
KS	1%	0%
KY	1%	0%
LA	8%	2%
MI	1%	0%
MO	20%	3%
NC	1%	0%
NJ	1%	0%

Table 3: Share of each state among efficient nursing homes and percentage of efficient nursing homes in each state

States	The share of each state among efficient nursing homes	Percentage of efficient nursing homes in each state
NV	4%	5%
OH	8%	1%
OK	9%	3%
SC	1%	1%
TX	15%	1%
WI	1%	0%

Then, by running the two-step cluster analysis for the 14073 NHs, it has been realized that by considering three clusters, we could obtain a high cluster quality based on the Silhouette measure of cohesion and separation (Figure 2). Figure 3 and Table 4 illustrate Schwarz' Bayesian Criterion BIC for the two-step cluster analysis, and Figure 4 displays a low interpretability of clusters based on the type and scale of the problem. In other words, the type and scale of the problem are important features for choosing the right number of clusters, and this importance can be interpreted by choosing a different number of clusters and observing the range of values in each cluster on the box plot, as shown in Figure 5. For three cluster, the silhouette measure of cohesion and separation is between the range of 0.5 and 1 (good range), also Auto clustering Schwarz's Bayesian Criterion (BIC) curve displays three cluster is the elbow point, and after that by the number of clusters increase there would be subtle change for BIC value. Moreover, the box plot in Fig 5 shows the efficiency range between the three cluster, without any overlap, it can be inferred that, we would face three distinct performances for the NHs. In other words, cluster 2 outperforms others (red line in Fig 5, and cluster 3 is at the middle (dark blue line) and its performance is greater than the third quartile, and cluster 1efficincy range (blue) is almost between first quartile and median, and Fig 6 illustrates that the majority of the NHs are in cluster 1 and further interpretations and results about the performance and characteristics of the clusters would be presented in the following.

Figure 2: Silhouette measure of cohesion and separation as cluster quality indicator

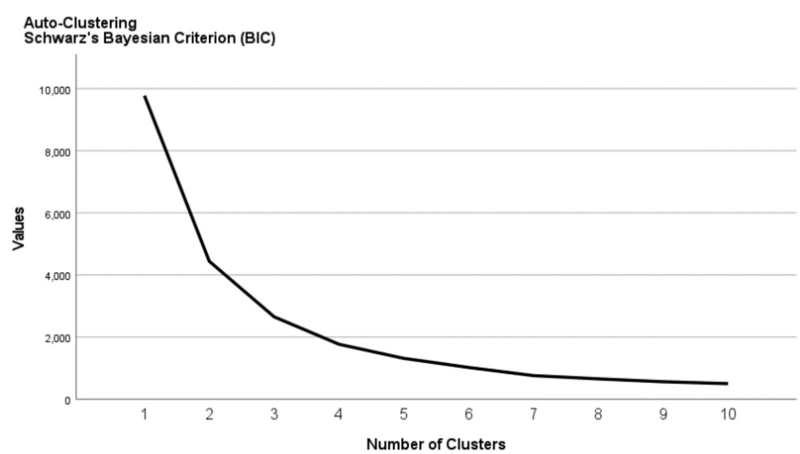


Figure 3: Auto clustering Schwarz's Bayesian Criterion (BIC) curve

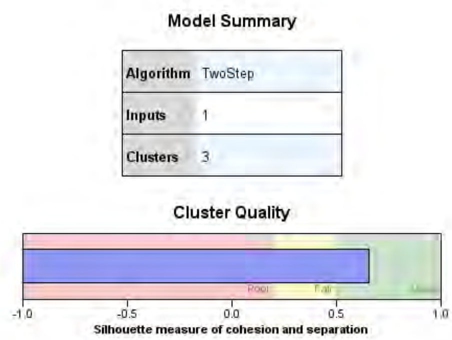


Figure 4: Auto clustering Ratio of Schwarz's Bayesian Criterion (BIC) change curve

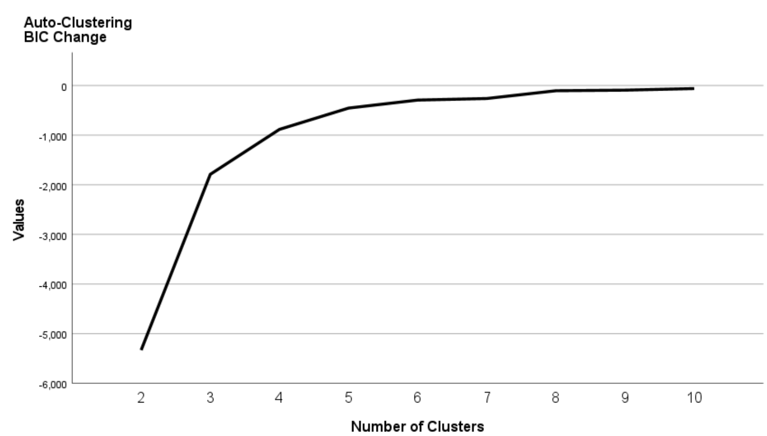


Figure 6 shows the size (percentage) of each cluster of NHs. Cluster 1 contains approximately 68.8% (9688 NHs), cluster 2 contains 4.6% (646 NHs), and cluster 3 contains 26.6% (3739 NHs) of the 14073 NHs.

Figure 5: Box plot for three clusters of nursing homes' efficiency scores

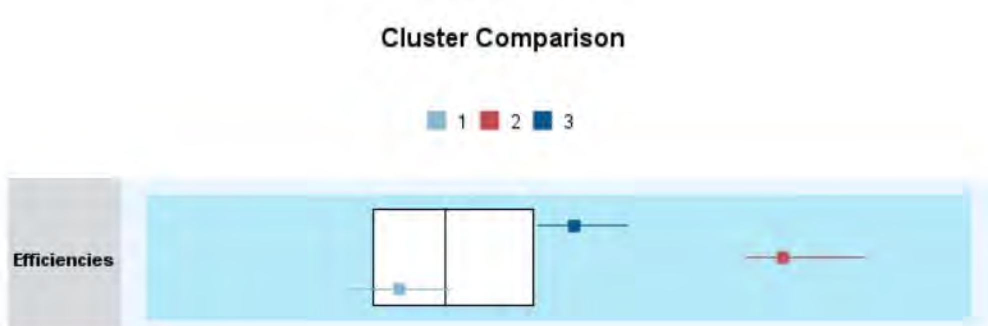
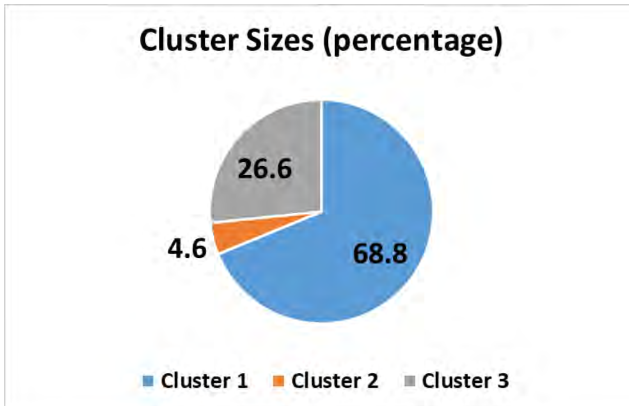


Figure 6: Sizes (percentage) of the clusters of NHs



Furthermore, the distribution of efficiency scores for each cluster can be used to show their difference and explain about tagging each cluster by a qualitative label based on their performance. Figures 7 ,8, and 9, respectively, display the efficiency score distributions for clusters 1, 2, and 3.

Table 4: Auto clustering Schwarz's Bayesian Criterion (BIC)

Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	BIC Change (a)	Ratio of BIC Changes (b)	Ratio of Distance Measures(c)
1	9773.264			
2	4442.828	-5330.436	1.000	2.959
3	2653.950	-1788.878	.336	2.003
4	1770.431	-883.519	.166	1.907
5	1316.148	-454.284	.085	1.517
6	1023.265	-292.883	.055	1.112
7	761.925	-261.340	.049	2.265
8	657.187	-104.738	.020	1.107
9	564.434	-92.753	.017	1.401
10	503.673	-60.760	.011	1.291

a. The changes are from the previous number of clusters in the table.

b. The ratios of changes are relative to the change for the two-cluster solution.

c. The ratios of distance measures are based on the current number of clusters against the previous number of clusters.

Figure 7: Efficiency score distribution of cluster 1

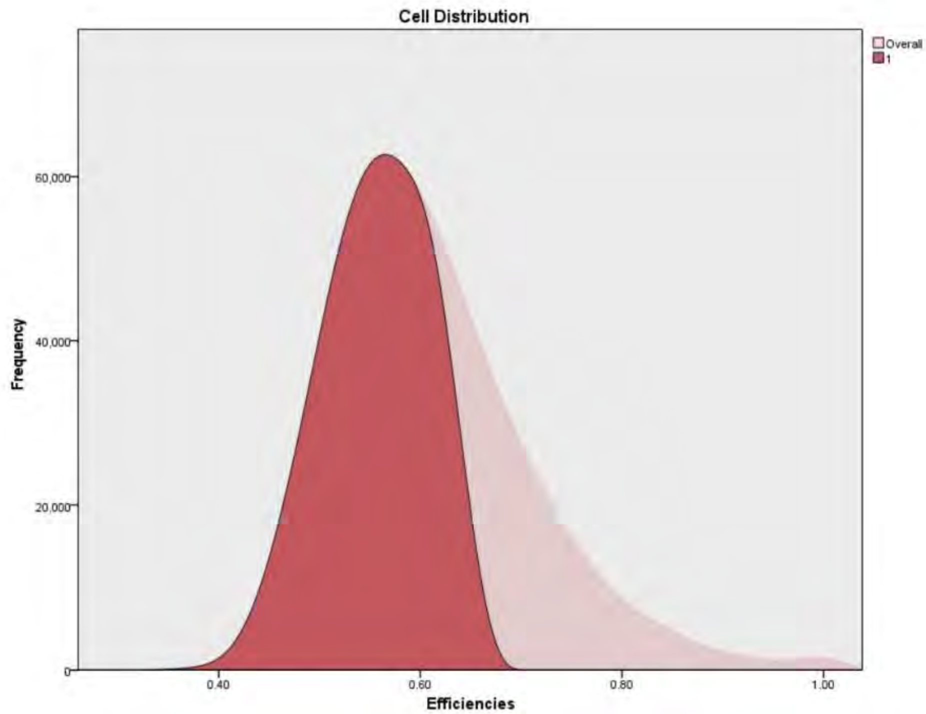


Figure 8: Efficiency score distribution of cluster 3

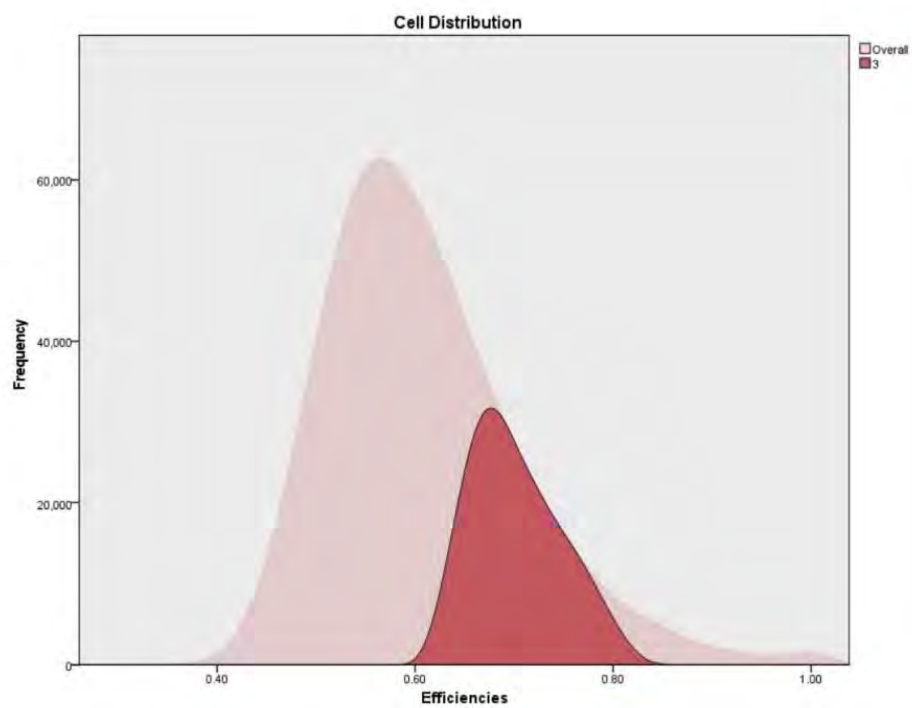
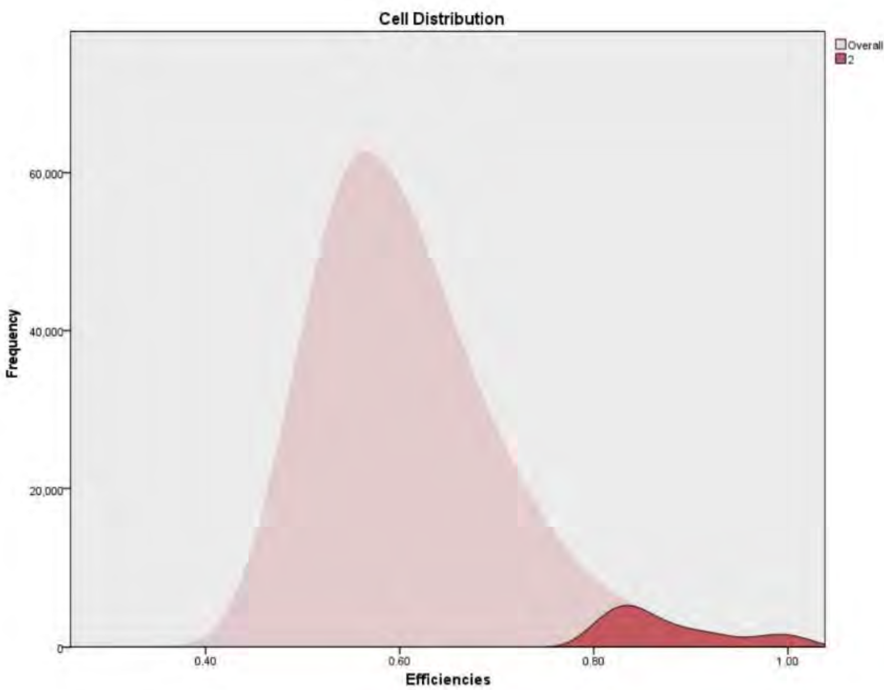


Figure 9: Efficiency score distribution of cluster 2



As previously stated, cluster 2 contains the fewest NHs (4.6%). Also, Figure 8 shows that the efficiency score of a NH in this cluster is usually between 0.8 and 1. Furthermore, all the 74 efficient NHs are included in this cluster. Thus, we refer to it as the *high-performance cluster* during COVID-19. Cluster 1 contains most of the NHs (68.8%), and the efficiency scores in this cluster range mostly between 0.4 and 0.6. Therefore, this cluster has been labeled as the *majority cluster*. Finally, cluster 3 with an efficiency range of nearly 0.6 to 0.8 has been labeled as the *medium-performance cluster*. Table 5 shows the proportion of NHs in each cluster with respect to each state.

Table 5: Share of each state's nursing homes in each cluster based on their performance in 2021 during COVID 19

States	cluster 1 Majority	cluster 2 High performance	cluster 3 medium performance	share of cluster 2 from efficient NHs in each state
AK	100%	0%	0%	
AL	77%	1%	22%	
AR	56%	4%	40%	
AZ	69%	1%	30%	
CA	57%	4%	39%	0.2%
CO	82%	2%	16%	0.02%
CT	95%	0%	5%	
DC	47%	7%	47%	
DE	98%	0%	2%	
FL	77%	2%	22%	0.1%

Table 5: Share of each state's nursing homes in each cluster based on their performance in 2021 during COVID 19

States	cluster 1 Majority	cluster 2 High performance	cluster 3 medium performance	share of cluster 2 from efficient NHs in each state
GA	64%	4%	32%	0.04%
GU	100%	0%	0%	
HI	96%	0%	4%	
IA	85%	1%	14%	
ID	75%	0%	25%	
IL	45%	14%	41%	1.12%
IN	68%	4%	28%	0.16%
KS	83%	1%	16%	0.01%
KY	83%	1%	16%	0.01%
LA	25%	21%	54%	1.68%
MA	84%	6%	11%	
MD	81%	4%	15%	
ME	96%	0%	4%	
MI	85%	1%	14%	0.01%
MN	94%	0%	6%	
MO	36%	18%	45%	3.6%
MS	82%	1%	17%	
MT	55%	3%	42%	
NC	65%	4%	31%	0.04%
ND	91%	0%	9%	
NE	84%	1%	16%	
NH	90%	0%	10%	
NJ	78%	1%	21%	0.01%
NM	57%	3%	40%	
NV	75%	6%	19%	0.24%
NY	91%	0%	9%	
OH	72%	3%	26%	0.24%
OK	37%	11%	52%	0.99%
OR	92%	1%	7%	
PA	77%	1%	22%	
PR	50%	25%	25%	
RI	77%	0%	23%	
SC	72%	3%	25%	0.03%
SD	92%	1%	7%	
TN	63%	5%	32%	
TX	34%	13%	53%	1.95%
UT	78%	2%	20%	
VA	82%	0%	18%	
VT	94%	0%	6%	
WA	91%	1%	8%	
WI	71%	2%	27%	0.02%

Table 5: Share of each state's nursing homes in each cluster based on their performance in 2021 during COVID 19

States	cluster 1 Majority	cluster 2 High performance	cluster 3 medium performance	share of cluster 2 from efficient NHs in each state
WV	93%	0%	7%	
WY	56%	0%	44%	

For cluster 2, 41 states have had NHs in the high-performance cluster. Considering Table 3, it should be noted that only 19 of these 41 states have efficient NHs (i.e., with an efficiency score of 1), and the remaining 22 states (AL, AR, AZ, DC, IA, MA, MD, MS, MT, NM, OR, PA, and PR), have NHs with efficiency scores ranging from 0.8 to less than 1. Another interesting finding is that the states with the highest proportion of NHs in the high-performance cluster include PR, LA, MO, IL, TX, and OK, with 25%, 21%, 18%, 14%, 13%, and 11%, respectively; however, PR is not in Table 3, i.e., it does not contain any efficient NHs, despite having the highest proportion of high-performance NHs. LA has 8% efficient NHs (Table 3), and 21% of its NHs are also in the high-performance cluster (Table 4). MO has the highest share of efficient NHs (20%), 18% of its NHs are in the high-performance cluster, and 3.6% of its high-performance NHs are efficient.

Table 5 also presents that all the states in the U.S. have NHs in the majority cluster with efficiency scores ranging from 0 to 0.4. All the NHs of Alaska (AK) and of Guam (GU) are in the majority cluster. Also, the average percentage of NHs in cluster 1 over all the states is 75%, which is why this cluster has been labeled as the majority cluster. Louisiana (LA) has the smallest percentage of NHs (25%) in this cluster, followed by TX (34%), MO (36%), and OK (37%).

Furthermore, cluster 3 is associated with medium-performance NHs with efficiency scores ranging from 0.6 to 0.8. Except Alaska and Guam, all the other states have a share in this cluster. On average, each state has a 22% of NHs in cluster 3, however the states with the largest shares are for LA (54%), followed by TX (53%) and OK (52%).

DISCUSSION AND CONCLUSIONS

Various research publications have discussed the benefits and capabilities of analytics for improving the healthcare sector (Kamble et al., 2018; Kumar et al., 2020; Wang & Hajli, 2017). Furthermore, in the face of a pandemic catastrophe and greater uncertainty, the uses of analytics have become more critical and lifesaving (Bag et al., 2021; Tortorella et al., 2022). The advantage of analytics applications in NHs is the same. The findings of this study inform about the performance of U.S. NHs during the epidemic. Concerning this, the proposed analytics approach and results in this study are useful for policy makers in this field.

From the risk management perspective, policy makers may recognize that most of the NHs (68.8%) are in *majority cluster* and the efficiency range for this cluster is between 0.4 to 0.6 (Figure 7). This might be viewed that many NHs have been facing serious challenges in the face of a pandemic. Furthermore, the proportion of each state in each cluster would disclose which states had more challenges than other states.

For more discussion and conceptualizing for expanding the results of this study to other relevant topics for NHs during COVID 19; the impact of other factors on the efficiency of NHs may be analyzed, such as vaccination attitudes and vaccine hesitancy in each state (Ozdenerol & Seboly, 2022; Pogue et al., 2020), ecological factors (Wu et al., 2020), longitudinal data of lifestyle of each state's residents until COIVD-19 (Arena et al., 2022), socioeconomic status of each state, and demographic trends (DuPre et al., 2021).

For instance, the rate of cardiovascular disease deaths is high in the Southeast region of the U.S. Also, there is a relationship between unhealthy lifestyle and the cardiovascular disease. On the other hand, evidence reveals the higher rate of hospitalization in these states (Arena et al., 2022). Also, our study shows that out of 14 states in Southeast region, only Kentucky, Louisiana, Missouri, North Carolina, and South Carolina could achieve a proportion in the group of states with the efficient nursing homes. However, Table 5 displays that for Louisiana and Missouri about 80% of their NHs are in *medium-performance cluster* and *majority cluster*, and for Kentucky, North Carolina, and South Carolina more than 90% of their NHs are in medium-performance cluster and majority cluster. Thus, it should be mentioned that one of our outputs has been average of the total residents COVID-19 deaths; although, the proportion of the NHs' residents who had been suffered from cardiovascular or other chronic diseases, has not been determined in this study, our results also indicate an improvement gap for those states. Due to the age of the NHs' residents, for involving the lifestyle of each state in the NHs' examination, a relevant time series dataset of each state's residents' lifestyle is required and by that we may involve a considerable interval of the resident's lifestyle for comparing it with the states' efficiencies.

Furthermore, this research provides valuable insights for resource allocation in the NH industry. With limited resources, policy makers and managers at a state level may prioritize their resources. In other words, they may see which NHs need to be more resilient to avoid the potential negative consequences of the pandemic crisis in the future.

NHs need different resources to have successful operations, such as clothing, personal care items, beds, bed linens, electronics and entertainment, hobby supplies, sentimental items, and decorations as well as some basic medical treatment equipment. Additionally, psychological factors and mental health have been determined as impactful factors for dealing with COVID-19 (Schulze et al., 2022), anxiety, stress, depression, sleep disturbances, death anxiety have been increased during the COVID-19 and overemphasis on the technical and tangible resources has resulted in the perceived erosion of care and stress on NHs' employees (Leontjevas et al., 2020). Also, higher mortality rate during pandemic has increased the psychological distress in the NHs and Because of the social isolation produced by the lengthy confinement, elderly persons in NHs report more psychological anguish and lower life satisfaction (Arpacıoğlu et al., 2021). Therefore, there has been psychological risk for both residents and staff in NHs. By referring to the proposed results in this study, the states with lower efficiency scores may be more susceptible to such distresses. Hence, policy makers may consider more attention for psychological and mental health problems of NHs in such states. Studies have suggested that during pandemic older individuals should be enabled to use telecommunication technology to communicate with family and friends, as well as to participate in person-centered activity programs that are convenient and safe (Arpacıoğlu et al., 2021). Hence, more computer literacy programs should be introduced to them as well as appropriate telecommunication and electronic devices. Also, recreational services (Kaelen et al., 2021); psychological consultations (Y. J. Kim et al., 2020); are of great importance and the state's governments may provide such resources and services for them.

To sum up, the proposed analytics application has presented insights about the efficiency of the NHs from the perspective of their performance against COVID-19 in each state of the U.S.

In this study, an analytics approach has been used to uncover information in the NH sector with the purpose of offering insights for better decision making. To our best knowledge, no research has directly addressed the performance of NHs during COVID-19 by analyzing measures on pandemic crisis management. This research examined the performance of 14073 NHs in the U.S. in 2021, and the data was extracted from the CMS. Our results illustrate the condition of NHs which is insightful for decision makers. Also, as the methodologies, DEA and two-step cluster analysis have been chosen. In this regard, the efficiency of each NH has been estimated by DEA, and then NHs have been clustered by two-step cluster analysis. Based on the average efficiency scores of each cluster which is a representative of their performance, the clusters have been labeled as *majority, medium-performance, and high-performance.*

This research is focused only on the specific indicators of the efficiency of NHs during COVID-19 and not the general efficiency; therefore, from the perspective of this research, 74 out of 14073 NHs have become efficient. Furthermore, out findings reveal 9688 NHs are in the majority cluster (68.8%); the high-performance cluster contains 4.6% of the NHs (646), and the medium-performance cluster has 26.6% of the NHs (3739). It demonstrates that 4.6% of the U.S. NHs could take appropriate measures in the wake of a pandemic outbreak. The other finding is that 19 out of 53 states and territories in the U.S. have efficient NHs. Having 15 efficient NHs out of 74, Missouri (MO) could become the most efficient state, followed by Texas (15%) and Oklahoma (9%) as the second and third efficient states.

The presented results in this research are comprehensive, and it adopted a national scope. Hence, it evaluates the efficiency of NHs by first considering each individual NH as the DMU (micro-level), and then by considering the NHs in each state, it presents the results at the state-by-state level. This approach would be useful for other researchers and macro-level decision makers in this sector.

As a future research suggestion, an in-depth examination of efficient NHs would be recommended for conceptualizing about lessons learned and recognizing the best practices, because these NHs had better resilience during the COVID-19. As a result, the lessons or practices are important for other NHs. Moreover, decision makers may pay attention to the states in the *high-performance cluster* as well as the mentioned states in the group of efficient NHs, demonstrated in Table 1. These states can be used as benchmarks, and the strategies and practices used in these NHs can internally be examined by more in-depth interviews as well as experiments. Additionally, the same analysis may be conducted to the states with the highest share in the high-performance cluster, and then researchers or decision makers may conduct more in-depth examination to understand their competitive advantages in this sector. Furthermore, psychological and behavioral factors play an important role in NHs. Therefore, by considering the performance of each state, researchers may conduct an experimental design to delve into the soft and behavioral issues of various NHs in different states to determine the role of factors that make NHs more efficient in terms of facing pandemic crisis. These lessons would be important for resisting future fatalities.

REFERENCES

Abrams, H. R., Loomer, L., Gandhi, A., & Grabowski, D. C. (2020). Characteristics of U.S. Nursing Homes with COVID-19 Cases. Journal of the American Geriatrics Society, 68(8), 1653–1656.

Arena, R., Pronk, N. P., Laddu, D., Whitsel, L. P., Sallis, J. F., & Lavie, C. J. (2022). Mapping One Million COVID-19 Deaths and Unhealthy Lifestyle Behaviors in the United States: Recognizing the Syndemic Pattern and Taking Action. The American Journal of Medicine, 135(11), 1288–1295.

Arpacıoğlu, S., Yalçın, M., Türkmenoğlu, F., Ünübol, B., & Çelebi Çakıroğlu, O. (2021). Mental health and factors related to life satisfaction in nursing home and community-dwelling older adults during COVID-19 pandemic in Turkey. Psychogeriatrics, 21(6), 881–891.

Bag, S., Gupta, S., Choi, T. M., & Kumar, A. (2021). Roles of Innovation Leadership on Using Big Data Analytics to Establish Resilient Healthcare Supply Chains to Combat the COVID-19 Pandemic: A Multimethodological Study. IEEE Transactions on Engineering Management.

Björkgren, M. A., Häkkinen, U., & Linna, M. (2001). Measuring efficiency of long-term care units in Finland. Health Care Management Science, 4(3), 193–200.

Chu, T., Society, J. S.-J. of the K. C., & 2021, undefined. (2021). Efficiency evaluation of nursing homes in China's eastern areas Based on DEA-Malmquist Model. Koreascience. 12(7), 273–282.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). Data envelopment analysis: History, models, and interpretations. International Series in Operations Research and Management Science, 164, 1–39.

Davidson, P. M., & Szanton, S. L. (2020). Nursing homes and COVID-19: We can and should do better. Journal of Clinical Nursing, 29(15–16), 2758.
de LUCA, R., & Calabrò, R. S. (2020). How the COVID-19 Pandemic is Changing Mental Health Disease Management: The Growing Need of Telecounseling in Italy. Innovations in Clinical Neuroscience, 17(4–6).

de Vries, J. M. A., & Curtis, E. A. (2019). Nursing leadership in Ireland: experiences and obstacles. Leadership in Health Services, 32(3), 348–363.

Dervaux, B., Leleu, H., Nogues, H., & Valdmanis, V. (2006). Assessing French nursing home efficiency: An indirect approach via budget-constrained DEA models. Socio-Economic Planning Sciences, 40(1), 70–91.

DuPre, N. C., Karimi, S., Zhang, C. H., Blair, L., Gupta, A., Alharbi, L. M. A., Alluhibi, M., Mitra, R., McKinney, W. P., & Little, B. (2021). County-level demographic, social, economic, and lifestyle correlates of COVID-19 infection and death trajectories during the first wave of the pandemic in the United States. Science of The Total Environment, 786, 147495.

Garavaglia, G., Lettieri, E., Agasisti, T., & Lopez, S. (2011). Efficiency and quality of care in nursing homes: An Italian case study. Health Care Management Science, 14(1), 22–35.

Gharoie Ahangar, R., Pavur, R., Fathi, M., & Shaik, A. (2020). Estimation and demographic analysis of COVID-19 infections with respect to weather factors in Europe. Journal of Business Analytics, 3(2), 93–106.

Hill, J. W., & Powell, P. (2009). The national healthcare crisis: Is eHealth a key solution? Business Horizons, 52(3), 265–277.

Iparraguirre, J. L., & Ma, R. (2015). Efficiency in the provision of social care for older people. A three-stage Data Envelopment Analysis using self-reported quality of life. Socio-Economic Planning Sciences, 49, 33–46.

Jalali Sepehr, M., Haeri, A., & Ghousi, R. (2019). A cross-country evaluation of energy efficiency from the sustainable development perspective. International Journal of Energy Sector Management, 13(4), 991–1019.

Kaelen, S., van den Boogaard, W., Pellecchia, U., Spiers, S., de Cramer, C., Demaegd, G., Fouqueray, E., van den Bergh, R., Goublomme, S., Decroo, T., Quinet, M., van Hoof, E., & Draguez, B. (2021). How to bring residents' psychosocial well-being to the heart of the fight against Covid-19 in Belgian nursing homes—A qualitative study. PLOS ONE, 16(3).

Kamble, S. S., Gunasekaran, A., Goswami, M., & Manda, J. (2018). A systematic perspective on the applications of big data analytics in healthcare management.
Kent, P., Jensen, R. K., & Kongsted, A. (2014). A comparison of three clustering methods for finding subgroups in MRI, SMS or clinical data: SPSS TwoStep Cluster analysis, Latent Gold and SNOB. BMC Medical Research Methodology, 14(1), 1–14.

Kim, N.-H. ;, Lee, J.-M. ;, Yoo, E., Tchounwou, B., Lazzer, S., Kim, N.-H., Lee, J.-M., & Yoo, E. (2022). How the COVID-19 Pandemic Has Changed Adolescent Health: Physical Activity, Sleep, Obesity, and Mental Health. International Journal of Environmental Research and Public Health 2022, Vol. 19, Page 9224, 19(15), 9224.

Kim, Y. J., Cho, J. H., & Park, Y. J. (2020). Leisure Sports Participants' Engagement in Preventive Health Behaviors and Their Experience of Constraints on Performing Leisure Activities During the COVID-19 Pandemic. Frontiers in Psychology, 11, 3392.

Kleinsorge, I. K., & Karney, D. F. (1992). Management of nursing homes using data envelopment analysis. Socio-Economic Planning Sciences, 26(1), 57–71.

Kumar, Y., Sood, K., Kaul, S., & Vasuja, R. (2020). Big Data Analytics and Its Benefits in Healthcare. Studies in Big Data, 66, 3–21.

Kuwahara, Y., Nagata, S., Taguchi, A., Naruse, T., Kawaguchi, H., & Murashima, S. (2013). Measuring the efficiencies of visiting nurse service agencies using data envelopment analysis. Health Care Management Science, 16(3), 228–235.

Leontjevas, R., Knippenberg, I. A. H., Smalbrugge, M., Plouvier, A. O. A., Teunisse, S., Bakker, C., Koopmans, R. T. C. M., & Gerritsen, D. L. (2020). Challenging behavior of nursing home residents during COVID-19 measures in the Netherlands.

Li, Y., Liu, A. C., Yu, Y. Y., Zhang, Y., Zhan, Y., & Lin, W. C. (2022). Bootstrapped DEA and Clustering Analysis of Eco-Efficiency in China’s Hotel Industry. Sustainability 2022, Vol. 14, Page 2925, 14(5), 2925.

Mangialavori, S., Riva, F., Froldi, M., Carabelli, S., Caimi, B., Rossi, P., Delle Fave, A., & Calicchio, G. (2022). Psychological distress and resilience among italian healthcare workers of geriatric services during the COVID-19 pandemic. Geriatric Nursing, 46, 132–136.

Mardani, A., Zavadskas, E. K., Streimikiene, D., Jusoh, A., & Khoshnoudi, M. (2017). A comprehensive review of data envelopment analysis (DEA) approach in energy efficiency. Renewable and Sustainable Energy Reviews, 70, 1298–1322.

Min, A., Park, C. G., & Scott, L. D. (2016). Evaluating Technical Efficiency of Nursing Care Using Data Envelopment Analysis and Multilevel Modeling. Western Journal of Nursing Research, 38(11), 1489–1508.

Ni Luasa, S., Dineen, D., & Zieba, M. (2018). Technical and scale efficiency in public and private Irish nursing homes – a bootstrap DEA approach. Health Care Management Science, 21(3), 326–347.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. International Journal of Surgery, 78, 185–193.

Ozdenerol, E., & Seboly, J. (2022). The Effects of Lifestyle on COVID-19 Vaccine Hesitancy in the United States: An Analysis of Market Segmentation. International Journal of Environmental Research and Public Health, 19(13), 7732.

Pogue, K., Jensen, J. L., Stancil, C. K., Ferguson, D. G., Hughes, S. J., Mello, E. J., Burgess, R., Berges, B. K., Quaye, A., & Poole, B. D. (2020). Influences on Attitudes Regarding Potential COVID-19 Vaccination in the United States. Vaccines 2020, Vol. 8, Page 582, 8(4), 582.

Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., & Parkinson, J. (2015). Using two-step cluster analysis to identify homogeneous physical activity groups. Marketing Intelligence and Planning, 33(4), 522–537.

Sajal Chattopadhyay, & Dennis Heffley. (1994). Are for-profit nursing homes more efficient? Data envelopment analysis with a case-mix constraint. JSTOR, 171–186. https://www.jstor.org/stable/40325557

Schulze, S., Merz, S., Thier, A., Tallarek, M., König, F., Uhlenbrock, G., Nübling, M., Lincke, H. J., Rapp, M. A., Spallek, J., & Holmberg, C. (2022). Psychosocial burden in nurses working in nursing homes during the Covid-19 pandemic: a cross-sectional study with quantitative and qualitative data. BMC Health Services Research, 22(1), 1–13.

Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. European Journal of Operational Research, 142(1), 16–20.

Shimshak, D. G., Lenard, M. L., & Klimberg, R. K. (2009). Incorporating quality into data envelopment analysis of nursing home performance: A case study. Omega, 37(3), 672–685.

Singh, K., Jahnke, I., Mosa, A., & Calyam, P. (2021). The Winding Road of Requesting Healthcare Data for Analytics Purposes: Using the One-Interview Mental Model Method for Improving Services of Health Data Governance and Big Data Request Processes. Journal of Business Analytics.

Sinimole, K. R. (2012). Evaluation of the efficiency of national health systems of the members of World Health Organization. Leadership in Health Services, 25(2), 139–150.

Soto-Ferrari, M., Carrasco-Pena, A., & Prieto, D. (2022). AGGFORCLUS: A hybrid methodology integrating forecasting with clustering to assess mitigation plans and contagion risk in pandemic outbreaks: the COVID-19 Case Study. Journal of Business Analytics.

Tkaczynski, A. (2016). Segmentation using two-step cluster analysis. Segmentation in Social Marketing: Process, Methods and Application, 109–125.

Tortorella, G. L., Fogliatto, F. S., Saurin, T. A., Tonetto, L. M., & McFarlane, D. (2022). Contributions of Healthcare 4.0 digital applications to the resilience of healthcare organizations during the COVID-19 outbreak. Technovation, 111, 102379.

Tran, A., Nguyen, K. H., Gray, L., & Comans, T. (2019). A Systematic Literature Review of Efficiency Measurement in Nursing Homes. International Journal of Environmental Research and Public Health 2019, Vol. 16, Page 2186, 16(12), 2186.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70, 287–299.

Watts, T., Tucker, M., Gray, C., Lee, K., Modina, K., & Gray, Z. (2021). Lessons learned in preventing COVID-19 within a skilled nursing facility during the early pandemic. Geriatric Nursing, 42(6), 1388–1396.

Wu, X., Nethery, R. C., Sabath, M. B., Braun, D., & Dominici, F. (2020). Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Science Advances, 6(45).

Zhang, L., Han, Y., & Fang, Y. (2021). Service Efficiency of Nursing Homes Based on DEA and Tobit Model: a Longitudinal Study in China.

DECISION SCIENCES INSTITUTE

Predicting Technology Replacement Timing

Askar H. Choudhury
Illinois State University
Email: choudhury@ilstu.edu

Nathan S. Hartman
Illinois State University
Email: nshartm@ilstu.edu

Ted A. Coussens
Illinois State University
Email: tacouss@ilstu.edu

ABSTRACT

An organization's ability to perform predictive maintenance in technology upgrades is essential in today's fast-paced technological innovation environment. This paper investigates a predictive framework to predict organization-wide major replacement timing for the upgrade of technological resources. In this paper, we present a predictive maintenance methodology through intervention modeling. A leading indicator is utilized to prognosticate points in the time expected to precede technology/equipment upgrades. The proposed theoretical framework identifies a leading indicator to detect early signals for an upgrade. Results show that "phone request" as a leading indicator signals the organization-wide major technology changeover/ replacement.

KEYWORDS: technology replacement, leading indicator, piecewise regression, structural break, predictive analytics

INTRODUCTION

The exponential growth in computer technology has resulted in abundant applications across all industries. The length of time a technology remains most valuable to an organization during its life span is uncertain. Therefore, identifying the most cost-effective time for an upgrade remains a challenge. Meanwhile, aging technologies require more repairs, force more downtime, and cause reductions in productivity (Marcinkowski & Gawin, 2019). Therefore, the motive for equipment/technology replacement in an organization is largely due to: 1) the likelihood of failure due to decay or high usage, 2) the sub-optimal performance of current equipment/technology because of the availability of higher-performing options (Bitros et al., 2010). The complexity of computer-intensive technologies has increased rapidly over the past three decades, making the timing for replacement more difficult. Research on organization-wide, large-scale replacement based on an aggregate group of users' features is absent in the extant literature. This omission has widespread consequences because most companies increasingly rely on computer technologies. This research is important for informing timely replacement decisions of technologies.

Organization-wide technology replacement/changeover differs from the replacement of individual units. Most of the existing literature is restricted to the narrower focus of individual equipment replacement timing (Goti et al., 2019). In this study, we examine company-wide equipment replacement timing for a group of units. Technology changeover is the large-scale replacement of an aggregate system adopted by an organization. In general, technology replacement decisions for the purpose of the upgrade are primarily user based, whereas replacement decisions for failure are primarily equipment/system based. Studies in the area of failure have focused mainly on single-unit systems (Kim & Makis, 2009; Delia & Rafael, 2008), and there is little research addressing multi-unit systems (Zhu et al., 2015; Salari & Makis, 2017). In the area of technology replacement for the upgrade at a micro-level, research has primarily focused on individual users' perspectives (Wang, Wang, & Lin, 2018). To our knowledge, only research at the macro-level, (Byun, et al., 2018) studied the technology upgrade timing for four separate industries using patent information as an indicator. Thus, at the macro level, questions regarding the timing for organization-wide equipment/technology replacement for the purpose of upgrade remain open. In this paper, we utilize a crucial leading indicator to predict replacement timing. Specifically, we apply piecewise regression for identifying when regime change will occur and to demonstrate how this is useful for making accurate predictions of replacement timing.

LITERATURE REVIEW

An individual component repair or replacement of a broken unit is indicative of traditional technology maintenance. However, a one-off event does not necessarily represent a situation where an organization requires a technology changeover. Changeover here represents the wholesale replacement of equipment/technology for the upgrade. A technology changeover is an upgrade of a technology system by replacing the majority of units on an enterprise-wide scale.

Operations managers often stress the need for implementing predictive maintenance of computer technology to prevent negative effects and maximize positive effects (Haynes & Thies, 1991). Predictive maintenance is also central to Industry 4.0, which promotes automation through computer technology and aims to achieve intelligent or smart factories (Goti et al., 2019; Lin et al., 2016). Recently, there has been a clear shift from corrective (re-active) maintenance (CM) strategies or even preventive (scheduled) maintenance (PM) to the more proactive predictive maintenance (PdM) strategy. Corrective maintenance activities are carried out at the point at which equipment/system failure occurs. Preventive maintenance is maintenance activities that are conducted at predetermined time intervals, and predictive maintenance anticipates the time at which early replacement can eliminate instances of equipment/technology failure. Predictive maintenance may utilize age-based and/or condition-based maintenance (CBM) approaches. Traditional age-based models primarily apply conventional Life Cycle Cost Analysis (LCCA) and the estimated age at which failure occurs may be in terms of time (years). Because LCCA focuses on future costs, assumptions are inevitable (e.g., the overall life cycle, repair and replacement cycles, discount rates, etc.). Thus, these assumptions should be carefully scrutinized to produce reasonable LCCA results. Conversely, condition-based maintenance (CBM) modeling predicts the equipment/technology failure timing based on data collected through system inspection via electronic sensors (Liao et al., 2012). Through this data, CBM models are able to calculate the deterioration or degradation levels of systems (Salari & Makis, 2017).

Technology manufacturers are continuously updating and developing new products, thereby diminishing the value of their earlier versions. As a result, organizations often need to replace their equipment/technology because they become obsolete. Moreover, new technologies have caused fundamental transformations in traditional business operations. A distinctive feature of the modern technology environment is the fast-paced innovations. This market characteristic, combined with the uncertainty of the usefulness of equipment/technology, presents a formidable challenge when organizations attempt to devise replacement strategies (Chronopoulos & Siddiqui, 2015). These factors put managers in situations where they are faced with the problem of timely replacement of technology with the appropriate versions of hardware and software.

Technology replacement timing for upgrade is the subject of extensive research focus at the micro (individual) level while understanding at the macro (organizational) level remains underdeveloped. The development of effective mechanisms of prediction at the macro level could have far-reaching significance for managers.

REFERENCES

References are available upon request.

DECISION SCIENCES INSTITUTE

Purposeful Internet Use and the Digital Divide in the United States: A Pandemic Outlook

Avijit Sarkar
University of Redlands
Email: avijit_sarkar@redlands.edu

James B. Pick
University of Redlands
Email: james_pick@redlands.edu

Owen Giron
University of Redlands
Email: owen_giron@redlands.edu

ABSTRACT

This paper analyzes geographic patterns and socioeconomic influences on purposeful internet use by American internet users during the COVID-19 pandemic. Purposeful internet use comprises online activities that range from social networking, teleworking, and accessing online education, entertainment, and financial services. K-means clusters show that the lowest levels of online activities take place in states that often had higher proportion of urban population and higher median age. Regression findings reveal that proportion of remote workers is positively associated with online activities, showing the pandemic's influence on purposeful internet use, a novel yet intuitive finding. Remote work is followed in importance by median age, urbanization, and educational attainment. Interestingly, urban population is found to be inversely associated with online activities. Implications of these findings for the digital divide in the U.S. are discussed and policy recommendations are outlined.

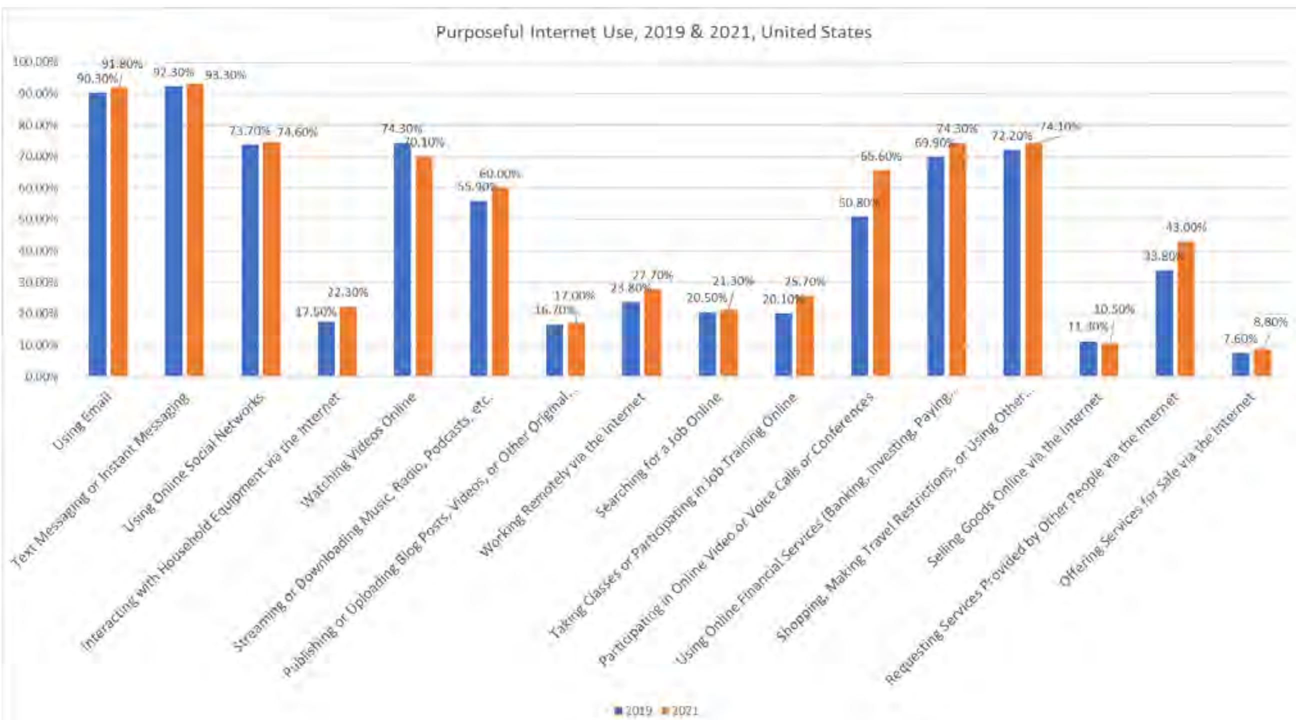
KEYWORDS: Online activities, Digital divide, Mapping, Cluster analysis, Regression

INTRODUCTION

Even prior to the COVID-19 pandemic, ubiquitous and affordable high-speed internet became an essential "utility" in American households to engage in a variety of online activities that affect daily life. For example, the internet is essential for electronic communication, online social networking, accessing employment opportunities and educational resources, as well as affordable healthcare options. Americans use the internet to listen to music, watch videos online, buy and sell goods online, offer and request services online, and interact with household devices connected to the internet.

During the COVID-19 pandemic, many of these purposeful uses of the internet – for e-communication, e-entertainment, e-education, e-health, e-commerce, and teleworking have expanded in scope and intensity. For example, prior to the pandemic, in November 2019, 51 million Americans used the internet to work remotely, which is 23.8 percent of all internet users age 15 years and older. At this time, remote workers constituted almost a third of estimated 160 million employed Americans (NTIA, 2019). Remote workers expanded to number 60.5 million, which is 27.7 percent of all internet users age 15 years and older, in November 2021. In contrast, a little over 16 million Americans (7.6% in November 2019) offered services for sale such as home rentals on short-term rental platforms (Airbnb) and ridesharing on platforms such as Uber or Lyft using the internet. Internet users engaged in this online activity expanded to number over 19 million internet users (8.8% in November 2021). Figure 1 depicts changes in purposeful internet use across an array of online activities between 2019 and 2021. It is clear from Figure 1 that traditional activities such as emailing and text and instant messaging are reaching a point of saturation, while newer and more contemporary online activities such as offering services for sale and interacting with household devices online are comparatively less prevalent. Yet the proportion of American internet users engaging in such activities is poised to grow based on the increase in the base and proportion of users between 2019 to 2021. The pandemic's influence is also evident in shaping the purposeful use spectrum. Participating in online video or voice calls or conferences grew the most among the sixteen online activities, by almost 15%, during 2019-2021 followed by requesting services provided by other people via the internet and taking online classes or participating in online job training, which grew by 9.20% and 5.60% respectively over the same period. As noted earlier, remote working also increased, somewhat modestly compared to these online activities.

Figure 1: Fluctuation in 16 Online Activities in U.S. States, 2019-2021 (Data: NTIA, 2021)



By purposeful internet use, we mean online activities using an internet-connected device such as a computer, tablet, or cell phone that are influenced by a specific purpose, for example, to engage in online communication with friends, family, or co-workers, to work remotely or attend classes remotely, to engage in online banking or bill payment, or any other financial activity. Purposeful internet use is linked to the attitude of an internet user. For example, an internet user purchasing a product online may perceive convenience to be a primary benefit of online shopping. Such a user is usually task-oriented and rational and may decide whether or not to purchase a product efficiently and rationally. Another user may enjoy time spent online to browse different products, search for deals, watch product videos and animations while simultaneously listen to music, and as such, find online shopping (Scarpi, 2011) a pleasurable activity. Based on the consumer's and internet user's persona, purposeful internet use may be utilitarian in the former case versus hedonic in the latter.

Going back to remote working, disparities were observed between race and ethnic groups, as well as for different occupations (NTIA, 2019). Similar disparities were also observed for taking online classes and participating in job training online, based on age. While the expansion in the overall spectrum of internet use as well as in per capita usage across different types of online activities is encouraging, differences persist based on race/ethnicity, age, income, educational attainment, and place of residence (Warf, 2012). These are critical factors that have traditionally influenced the digital divide in the United States (NTIA, 2011). In fact, age-based disparities in internet access and use have persisted, and the urban-rural gaps in broadband access and usage have been enduring, highlighting the geographic nature of the digital divide in the United States. Against this backdrop of diversification of online activities and expansion in the base of users across myriad online activities, it is critical to focus on the geography of purposeful internet use in the U.S. and analyze its demographic, social, educational, economic, and occupational underpinnings to discern facilitators and barriers of this usage dimension of the digital divide.

The objective of this paper is to analyze geographic patterns and demographic and socioeconomic influences on purposeful internet use in U.S. states. The research questions are:

1. What are the geographic patterns and disparities in purposeful internet use for online activities spanning communication, entertainment, work and education, and financial and commercial online activities?
2. What are the geographic clusters of these categories of purposeful internet use in the U.S. and what are the demographic and socioeconomic characteristics of such clusters?
3. What are the associations of demographic, economic, occupational, urbanization, infrastructure, social capital, societal openness, and affordability on purposeful internet use in U.S. states?

This study is novel due to its focus on online activities in the United States during the COVID-19 pandemic. Specifically, the data used for this study are from November 2021, and sourced from Internet Use Survey of the National Telecommunications and Information Administration (NTIA). This survey collected the first comprehensive federal data on how internet use in America evolved since the onset of the COVID-19 pandemic. Hence this study is one of the first in shedding light on geographic dimensions and socioeconomic influences on purposeful internet use due to the COVID-19 pandemic.

Another notable aspect of this study is its focus on analyzing the geography of purposeful internet use during the COVID period. Mapping the digital divide has been a federal policy initiative of the Federal Communications Commission (FCC) since before the pandemic. An outcome is the National Broadband Map and various recent map-based dashboards such as the Access Broadband dashboard -- all stemming from the Access Broadband Act, enacted in 2021 with the goal of increasing access to high-speed internet by expanding broadband networks to communities in need. In this study, the per capita base of users engaged in various online activities is mapped at the U.S. state level. Geographic interactions with demographic and socioeconomic attributes are examined. By accounting for the possible presence of geographic bias due to proximity, for example, between neighboring states, the actual influence of race/ethnicity, income, urbanization, and other factors can be accurately estimated. Geographic mapping and analysis are also key to understanding inequities in purposeful internet use and the extent of the digital divide in the U.S.

This study is conducted at the state level and is influenced by the fact that data on online activities is readily available at the state level. While this is opportunistic and has methodological implications and limitations, U.S. states often shape their own telecommunications and internet policies, in concert with federal policies and regulations. The findings of this study can be used to inform and design federal- and state-level policies, determine infrastructure development, capacity reinforcement, and related grants and awards for distressed communities, with an eye on leveling the digital divide and ensuring progress towards digital equity.

The remainder of this paper is organized into section on literature review of broadband and internet use, particularly in the U.S., a conceptual model of purposeful internet use, methodology and data, geographic patterns of online activities, regression findings, implications, limitations, and directions for future research.

LITERATURE REVIEW

The digital divide has been defined as "the gap between individuals, households, businesses and geographic areas at different socioeconomic levels with regard both to their opportunities to access information and communication technologies (ICTs) and to their use of the Internet for a wide variety of activities (OECD, 2001, p.5)." This definition refers to the gap in terms of both access and use of ICTs such as the internet. The access gap, often referred to as the "first-level digital divide" has been extensively studied (Lythreatis, Singh, and El-Kassar 2022). A shift in focus from access to digital skills and digital usage is considered the "second-level digital divide". The second level digital divide has also received a lot of attention from researchers who have studied disparities in technical means, autonomy of use, patterns of use, and digital skills (Mossberger et al., 2003; Van Deursen and Van Dijk, 2011; Van Deursen et al., 2016). A shift in focus from digital skills and usage to beneficial outcomes of using the internet is referred to as the "third-level of the digital divide."

This paper focuses on the second level of the digital divide and its focus is on purposeful internet use in U.S. states. A previous study analyzed factors associated with ICT access and use in U.S. states. ICTs included in the study were desktop/laptop in the household, broadband adoption in the household, households that were cellphone only, and access to mobile-wireless high speed devices, plus social media (Facebook and Twitter) access. This study found that social capital, education, societal openness, urbanization, and ethnicities are significantly associated with ICT access and use in U.S. states (Pick, Sarkar, and Johnson, 2015). The study was notable for its validating that social connections had a positive influence on ICT access and use in U.S. states. Another study whose principal focus was to understand Internet usage patterns in the U.S. used individual and household level data to find that peer effects based on locational proximity are critical in explaining large variations in internet use (Agarwal, Animesh, and Prasad, 2009). The implications of social capital for internet use and online communication was studied by Chen (2013). This study used individual-level survey data and found that resource-rich bonding social capital helps overcome the digital divides in access, general use, and online communication. At the county level, an earlier study found that professional/scientific/technical services workforce, other services workforce, household income, federal grant funds, college education, and ethnicity were key determinants of IS data processing, receipts, and payroll, for a sample of 164 counties (Azari and Pick, 2005). More recently, another county level study spanning 3,108 counties examined social media access and use (Facebook, Twitter, and LinkedIn). This study found that socioeconomic factors including demography, economy, education, innovation, and social capital were posited to influence social media utilization. The study also documented disparities in social media access and use between metropolitan, micropolitan, and rural counties (Pick, Sarkar, and Rosales, 2019).

In terms of purposeful internet use, one the earliest studies found that the type of internet connectivity (dial-up versus broadband) played a key role in explaining differences in the extent that the internet was used in American households to search information online using a search engine, read online news, find information about jobs online, check weather-related updates, read or write an online blog, download a podcast, engage in social networking, play online games, and download or share files using a peer-to-peer network (Warschauer and Matuchniak 2010). A more recent study examined geographic patterns of purposeful use of mobile internet in the U.S. at the state level. Purposeful use variables included indicators of mobile internet use for personal (text friends and family, view the news, watch videos, and engage in social networking) and business (conduct mobile banking, bill payment, research products, make purchases, and redeem coupons) purposes. The study found that age structure, employment in management, business, science, and arts occupations, affordability, and the extent of freedom in US states were important in explaining variations in mobile internet purposeful use (Sarkar, Pick, and Bhat, 2022).

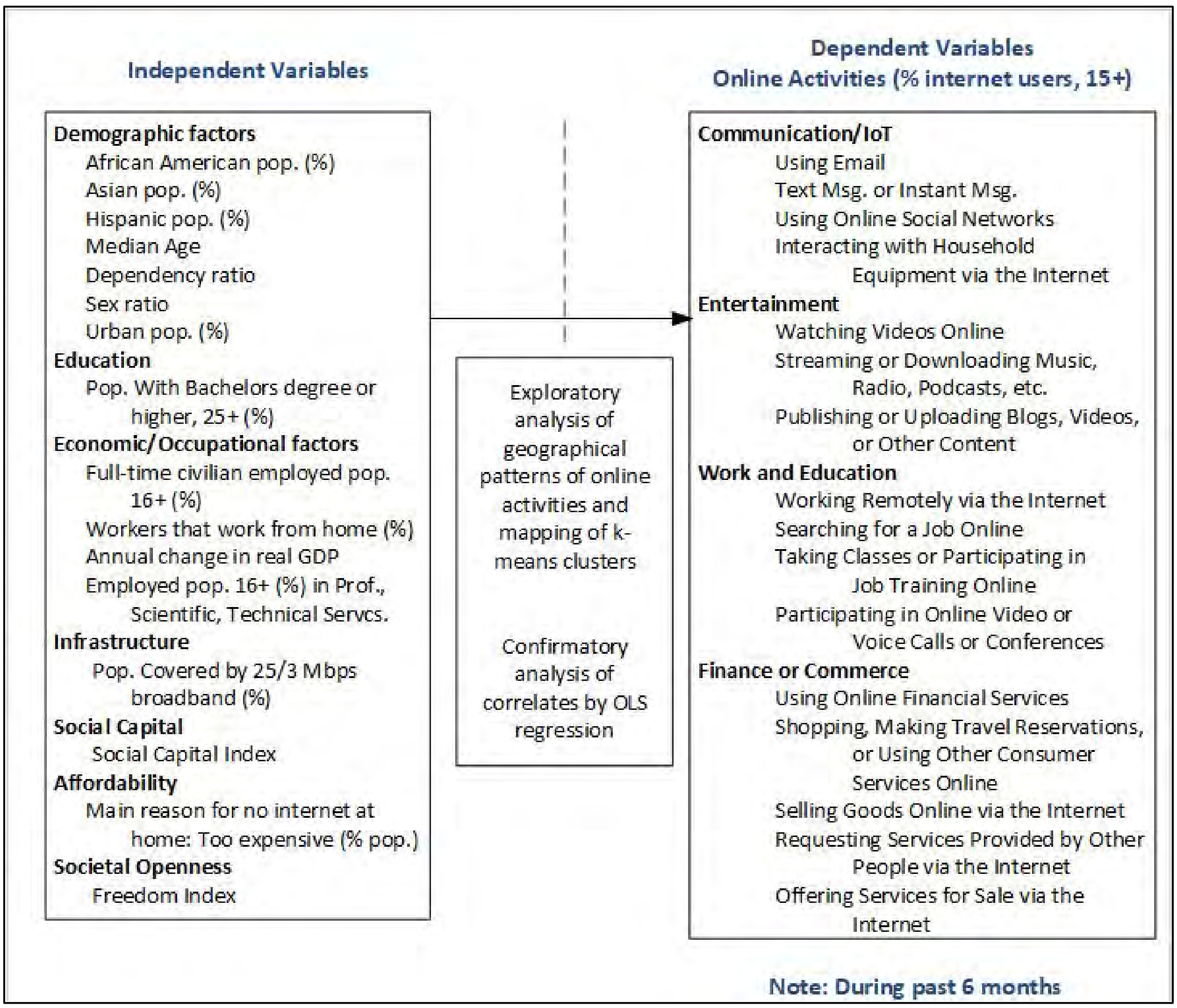
To the best of our knowledge, no prior study has examined purposeful use of the internet in the United States across a broad spectrum of online activities during the COVID-19 pandemic. Hence this is a unique feature of this study along with its focus on analysis of geographic patterns of such use in U.S. states.

CONCEPTUAL MODEL OF PURPOSEFUL INTERNET USE

Over the years, digital divide researchers have proposed a number of theories and conceptual models to examine various aspects of the digital divide. Well-known theoretical models include the Adoption-Diffusion Theory (Rogers, 2010), the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh et al., 2003), and the Resources and Appropriation Theory of the diffusion, acceptance, and adoption of new technologies (van Dijk, 2005). While a consensus theory or framework is yet to emerge in the digital divide literature, it is widely understood and theoretically and empirically validated that personal categorical inequalities (factors such as age, gender, race/ethnicity, intelligence, personality, and health), positional categorical inequalities (educational attainment, occupation, type of household, and nation – developed/developing, region – urban/rural), resources (time, money, social connections, mental abilities, motivation, and social status) often influence the access of a new digital technology and ultimately its use. Some of the theories are better suited to examine the digital divide at the level of individual adopters and users of information and communication technologies (ICTs).

In this study, our conceptual model of purposeful internet use shown in Figure 2 is based on the Spatially Aware Technology Utilization Model (SATUM) (Pick and Sarkar, 2015). SATUM is comprised of a group of dependent indicators of ICTs, in this case, online activities that constitute purposeful internet use, and a set of independent variables that influence the extent to which ICTs are accessed and used. A key aspect of SATUM is its spatial awareness. SATUM recognizes that geography and location plays a key role in the adoption, diffusion, and use of ICTs and it is essential to model the role of geography in empirically validating posited association between the set of independent variables and the set of dependent ICTs indicators. SATUM can be applied at various units of geography, for example, it has been used at the national level to analyze the digital divide in Africa (Pick and Sarkar, 2015). It has also been used to study the digital divide in the United States at its state (Pick, Sarkar, and Johnson, 2015) and county (Sarkar, Pick, and Rosales, 2023) levels. Modeling geographic patterns, agglomerations, and consequent bias in the adoption, diffusion, and use of ICTs is a distinguishing feature of SATUM. More information about SATUM can be found in Pick and Sarkar (2015).

Figure 2: Conceptual model of purposeful internet use



Our conceptual model of purposeful internet use for U.S. states (shown in Figure 1) is comprised of 16 independent variables, partitioned into 7 categories, that are posted to be associated with 16 dependent indicators of purposeful internet use, split into 4 categories. These indicators are online activities conducted by American internet users during and prior to the COVID-19 pandemic. All the independent variables and dependent indicators included in the conceptual model are either induced from prior literature (Stebbins, 2001) or authors' reasoning is provided for their inclusion. This justification is discussed next, first for the independent variables and then for the dependent indicators of internet use.

Justification for Independent Variables

There are 7 categories of independent variables – demographic, education, economic and occupational, infrastructure, social linkage, affordability, and societal openness.

Demographic Variables

The demographic variables included in the conceptual model span race/ethnicity, age, and gender. Among race/ethnic variables, broadband adoption and internet use has been found to be positively associated with Asian population, and often inversely with African American and Hispanic population in the United States (Pick, Sarkar, and Johnson, 2015; NTIA, 2013). For online activities such as social networking, parity in internet use has been achieved according to the latest NTIA data, yet disparities continue to remain for e-commerce activities such as shopping, making travel reservations, and obtaining other consumer services online (NTIA, 2021).

Age and gender are well-known factors influencing digital divides all over the world. Traditionally, higher age has been associated with lower ICT adoption, diffusion, and use, while females are less likely to be adopters and users of the internet than males, particularly, in developing regions. In this study, age and sex ratio are posted to be associated with the internet use indicators with age being inversely associated and sex ratio (males per 100 females) being positively associated. Children and the elderly dependents in the household are also likely to influence internet usage in various ways. During the pandemic, the presence of school-going children in the household is likely to have necessitated the use of internet for attending classes online, while the presence of the elderly is likely to decrease internet use. We posit that dependency ratio (number of dependents aged zero to 14 and over the age of 65, compared with the total population aged 15 to 64) will be associated with internet use but the direction of association can be mixed depending on the purpose of usage.

Finally, population size, density, and the degree of urbanization are also correlated with the digital divide (Vicente and Lopez 2011). Urban residents have been documented to be more frequent and heavier users of the internet than their rural counterparts. In 2021, the urban-rural gap in internet use in the U.S. was less than 2%. However, for certain types of online activities such as accessing entertainment and financial services online, this gap was more than 10% (NTIA, 2022). We posit proportion of urban population to be positively associated with internet use.

Education Variable

Educational attainment is another well-known factor influencing the digital divide. In a state-level study of ICT access, adoption, and use in U.S. states, proportion of college graduates was found to be positively associated with use of desktop/laptop, internet access, broadband adoption, and social networking (Pick, Sarkar, and Johnson, 2015). Those with a college degree are likely to be more skilled and likely more motivated at using the internet for a variety of purposes. We posit proportion of population age 25+ with bachelors degree or higher to be positively associated with internet use.

Economic and Occupational Variable

As discussed earlier, positional categorical inequalities such as employment and occupation are related to the digital divide. We posit civilian employed population (age 16+) and change in state GDP to be positively associated with internet use. Professional, scientific, and technical services employment was a key correlate of ICT adoption in a prior study of U.S. counties (Azari and Pick, 2005) and is posited to be positively associated with the dependent variables. We also reason that as more Americans worked from home during the COVID-19 pandemic, it is likely that overall internet use expanded and was in fact accelerated. This manifested across different purposes of internet use and as such proportion of remote workers is posited to be positively associated with the study's dependent variables. Median household income is excluded as an economic variable despite its well-known association with the digital divide. While screening for multicollinearity, we found median household income to be highly positively correlated (Pearson correlation coefficient of .824 significant at .001 level) with educational attainment (Bachelors degree or higher); hence it was excluded.

Infrastructure, Social Capital, Affordability, and Societal Openness

The deployment of broadband infrastructure to alleviate digital inequalities has been a key policy initiative as part of the recent Infrastructure Investment and Job Act of the U.S. Congress (U.S. Congress, 2021). Policy initiatives include mapping unserved and underserved locations. Unserved locations are identified as those lacking broadband service with a speed no less than 25 megabits (Mbps) per second for downloads, 3 megabits per second for uploads, and latency sufficient to support real-time, interactive applications. Since unserved communities and locations are unlikely and unable to use the internet, we posit proportion of population covered by 25/3 Mbps broadband to be positively associated with internet use. Prior studies of the United States found ICT access and broadband use to be positively associated with social capital for U.S. states (Chen 2013, Pick, Sarkar, and Johnson, 2015). Social chasms often adversely affect digital equity and social capital is known to provide digital access and enhance digital skills of less-skilled or unskilled digital users and nascent adopters of technologies such as IoT. In recent surveys of computer and internet use in the household, it has been repeatedly documented that affordability is the primary reason for non-adoption of the internet, followed by no reason to use, and privacy concerns (NTIA, 2021). Hence, we include this variable and posit it to be inversely associated with internet use. Finally, a more free and open society is more likely to encourage the exchange of ideas, information, data, and internet resources. Hence societal openness, measured by a Freedom Index, that accounts for personal, fiscal, regulatory, and economic freedom is posited to be positively associated with internet use.

Justification for Dependent Variables

Sixteen dependent indicators of purposeful internet use span four categories of dependent variables – online activities that relate to electronic (i) communication, (ii) entertainment, (iii) work and education, and (iv) financial and commercial activities. These dependent variables represent a broad array of online activities that require differing levels of motivation, skills, and resources (for example, devices such as computer, tablet, or cell phone) to engage in.

Communication

The indicators of purposeful internet use for communication are using email, text messaging and instant messaging, using online social networks, and communicating with household equipment using the internet. The first three indicators are traditional forms of communication between people using the internet, while the last indicator is a contemporary form of internet use for communicating and interacting with household equipment, devices, or appliances that are connected to the Internet, such as a connected thermostat, light bulb, or security system. It is to be noted that this category does not include communicating with digital assistants such as Apple’s Siri or Amazon's Alexa.

<u>Entertainment</u>

Entertainment is comprised of three indicators: watching videos online, streaming or downloading music, video, or podcasts, and publishing or uploading, blogs, videos, and other digital content online.

<u>Work and Education</u>

The four indicators of work and educational online activities include working remotely using the internet searching for jobs online, taking classes or participating in job training online, and participating in online video or voice calls and conferences using Zoom, Skype, FaceTime, and similar technologies.

<u>Financial and Commercial Activities</u>

The final category of online activities, financial and commercial include using the internet to access financial services (online backing, investing, bill payment), online shopping, making travel reservations, or purchasing other consumer services online, selling goods online (on platforms such as eBay, Etsy), requesting services provided by other people via the internet (for example, ride hailing, making short term rentals on platforms such as Airbnb, grocery and food delivery using platforms such as Instacart and DoorDash and requesting repair services using platforms such as Angie’s), and offering services for sale online (for example, ridesharing and offering rentals on platforms such as Uber, Lyft, and Airbnb).

We argue that disparities in engagement in these types of online activities particularly in light of the COVID-19 pandemic represents the new frontier of digital divide research. Hence, we include a broad spectrum of online activities that constitute purposeful use of the internet in an American household. Internet use to access healthcare resources, health information, insurance coverage information, telemedicine services are all part of the growing domain of electronic health. Also, accessing government services, such as registering to vote or renewing one’s driver’s license have been facilitated by using the internet. While omitted from this study, e-health and e-government related online activities would provide a more comprehensive and holistic view of American internet users and the usage digital divide. This is outlined as a next step of this study.

METHODOLOGY

The methodology of this study is comprised of multiple steps. The dataset was first compiled using a variety of sources that are described in the next section and verified for completeness and accuracy. Next, the variables were normalized as appropriate and descriptive statistics were computed.

In total, the dataset has sixteen dependent variables that are online activities conducted by Americans during the pandemic. The dataset originally had 19 independent variables, from which three were excluded for the purpose of developing regression models. This was due to screening independent variables for multicollinearity, using the Pearson correlation coefficient. One variable excluded is labor force participation rate, an economic indicator that was highly positively correlated with several other independent variables included in the conceptual model. An indicator for infrastructure – broadband subscribers in the household was also excluded for similar reasons. Finally, median household income was also excluded as an explanatory variable due to strong positive correlations with proportion of remote workers (Pearson Correlation Coefficient of .887 significant at the .001 level), professional, scientific, and technical services employment (Pearson Correlation Coefficient of .876 significant at .001 level), and educational attainment (Bachelors degree or higher). After screening for multicollinearity, sixteen independent variables were left for regression modeling.

The dependent variables were subsequently mapped using a Geographic Information System (GIS). Mapping the digital divide is useful to identify areas (states, provinces, or counties) with digital disparities. The geography of the digital divide been an increasing policy focus of the U.S. Federal Communications Commission (FCC). While mapping online activities in U.S. states, we overlaid independent variables of interest such as urban population to observe interactions between geographic patterns of online activities and the extent of urbanization (as one example) in U.S. states. Mapping while descriptive and exploratory can often provide important visual cues about clusters and outliers of online activities, which have implications for the diagnosis of spatial bias.

To diagnose spatial bias, spatial autocorrelation analysis was performed using the Moran's *I* test statistic (Moran, 1950), which is computed as follows:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}\, z_i\, z_j}{\sum_{i=1}^{n} z_i^2}$$

z_i is the deviation of an attribute for feature i from its mean ($x_i - \bar{x}$), $w_{i,j}$ is the spatial weight between features i and j, n is the total number of features, and S_0 is the sum of all spatial weights. Moran's *I* measures the extent of spatial autocorrelation of online activities in US states. The Moran's I test is inferential; the null hypothesis is that the values of a variable are randomly distributed spatially. The test statistic ranges in value between –1 and +1. Moran's I statistic value close to 0 for a dependent variable (online activity) would indicate spatial randomness while values close to –1 and +1 indicate the presence of spatial bias for a dependent variable that needs to be accounted for while examining associations of independent variables with the dependent variable in question. Interpretation of Moran's I is performed using the *p* value for statistical significance (if *p* is not significant, the variable is randomly distributed spatially). Further, if the Z score is positive, the values of a variable are more geographically agglomerated (high values located near high ones and low values near low ones). If it is negative, the spatial pattern resembles a "checkerboard" pattern, in which high values are surrounded by low ones and vice versa (Moran 1950, Openshaw, 1984).

K-means clustering, an unsupervised data mining process was next applied to partition the U.S. states into clusters based on their intensity of online activities for four different categories – communication, entertainment, work and education, and finance and commerce. Cluster centers were computed and the clusters were also mapped using a GIS to visualize geographic patterns of clusters. This provides further visual insights about the spatial agglomeration of online activities in U.S. states. Each cluster was characterized in terms of its demographic, social, economic, occupational, urbanization, infrastructure, and affordability attributes and similarities and differences were observed. While cluster characterization is descriptive, it provides clues about demographic and socioeconomic influences on the extent of online activities in U.S. states.

To analyze associations of demographic and socioeconomic variables on online activities, Ordinary Least Squares (OLS) regressions are conducted, in stepwise fashion, with significance levels of equal or less than 0.05. Two rounds of OLS regressions was conducted for this study, first with the independent variable proportion of remote workers included, and then excluded. This variable was found to have a dominant influence in the first round of OLS regressions. Hence, as a robustness test, remote workers was excluded to identify any additional variables that significantly influence the variation of online activities. The Variance Inflation Factor (VIF) was calculated for each regression to diagnose multicollinearity, and a VIF cutoff of 5.0 was used (Myers, 1990). No multicollinearity problems were detected in either round. Lastly, spatial autocorrelation of the OLS regression residuals were measured using Moran's I to identify the extent of randomness versus the extent of clustering of residuals. The residuals were also mapped to visualize their geographic patterns. Measuring the extent of agglomeration of residuals and mapping them provides further verification and validation of the robustness of the study's conceptual model.

DATA

Data for the study's sixteen dependent and sixteen independent variables were collected and then compiled from a variety of different sources.

As discussed earlier, the sixteen dependent indicators of internet use in this study span four categories of online activities – communication, entertainment, work and education, and finance and commercial activities. Broadly speaking, these sixteen indicators represent a broad spectrum of online activities that span traditional to more modern, contemporary uses of the internet. As shown in Figure 1, some of these online activities became more popular during the COVID-19 pandemic and the use of the internet for other activities were certainly catalyzed due to the pandemic.

Data for the sixteen dependent indicators of online activities in U.S. states were collected from the NTIA's Digital Nation Data Explorer (NTIA, 2021) for the period November 2021. The data published in the Digital Nation Data Explorer is compiled from the Computer and Internet Use Supplement that is part of that month's (November 2021) Current Population Survey, conducted by the U.S. Census Bureau (US Census Bureau, 2021). The Computer and Internet Use Supplement is sponsored by the NTIA and the survey data are useful to understand why, where, and how Americans use the internet and what are the barriers of effective internet use and consequent participation in the digital economy. As such, NTIA's Digital Nation Data Explorer is an authoritative source of data on purposeful internet use in the United States.

The independent variables represent demographic attributes that include race and ethnicity, age-related attributes such median age and dependency ratio, and sex ratio, educational attainment represented by proportion of population with a Bachelors degree or higher, and population density represented by proportion of urban population. Economic and occupational attributes include median household income, full-time civilian employed population age 16+, proportion of remote workers, annual change in state GDP, and employment in professional, scientific, and technical services (PSTS). Other independent variables represent infrastructure, social capital that is indicative of social interconnectedness, societal openness, and affordability. Data for the demographic, educational attainment, urbanization, and some of the economic variables were collected for U.S. states from the U.S. Census Bureau's 2017-2021 5-Year American Community Survey (ACS) Estimates. Data for annual change in real state GDP is sourced from the 2021 Regional Economic Accounts, Bureau of Economic Analysis.

For the infrastructure indicator that indicates population coverage by broadband fixed services of at least 25/3 Mbps, data were collected from the 14th Broadband Deployment Report of the Federal Communications Commission (FCC). Data on social capital in the U.S. states were obtained from a report commissioned by the Joint Economic Committee of the U.S. Congress (US Senate, 2018). The state-level index is comprised of four sub-indices that focus on family unity, family interaction, social support, community health, institutional health, collective efficacy, and philanthropic health. Additional details about the social capital index and its research design methodology can be found in the report (US Senate, 2018). Data on freedom in U.S. states for 2018 were obtained from the Cato Institute's Freedom in the 50 states project report (Ruger and Sorens, 2021). Finally, data for an independent variable that is an indicator of affordability were collected from the NTIA's Digital Nation Data Explorer.

All sixteen dependent variables and ten of the sixteen independent variables were normalized based upon 2017-2021 5-Year American Community Survey (ACS) population estimates in U.S. states and the District of Columbia. The study's codebook and descriptive statistics of all dependent and independent variables are in Table 1.

Table 1: Variable Definitions and Descriptive Statistics

Category of Dependent Var.	Online Activity (% of internet users, age 15+)	Minimum	Maximum	Mean	SD
Communication / IoT	Using Email	0.663	0.997	0.759	0.058
	Text Messaging or Instant Messaging	0.674	0.992	0.769	0.056
	Using Online Social Networks	0.482	0.814	0.615	0.056
	Interacting with Household Equipment via the Internet	0.103	0.254	0.175	0.034
Entertainment	Watching Videos Online	0.480	0.865	0.581	0.065
	Streaming or Downloading Music, Radio, Podcasts, etc.	0.399	0.795	0.498	0.064
	Publishing or Uploading Blog Posts, Videos, or Other Original Content	0.077	0.210	0.131	0.027
Work/Education	Working Remotely via the Internet	0.135	0.580	0.224	0.062
	Searching for a Job Online	0.092	0.312	0.166	0.034

	Taking Classes or Participating in Job Training Online	0.130	0.400	0.211	0.045
	Participating in Online Video or Voice Calls or Conferences	0.443	0.896	0.537	0.068
Financial / Commercial	Using Online Financial Services (Banking, Investing, Paying Bills, etc.)	0.490	0.883	0.616	0.066
	Shopping, Making Travel Reservations, or Using Other Consumer Services Online	0.453	0.880	0.621	0.069
	Selling Goods Online via the Internet	0.056	0.155	0.091	0.023
	Requesting Services Provided by Other People via the Internet	0.178	0.795	0.339	0.085
	Offering Services for Sale via the Internet	0.024	0.138	0.066	0.023
Category of Independent Var.	Independent Variable	Minimum	Maximum	Mean	SD
Demographic	Black or African American	0.005	0.432	0.108	0.102
	Asian	0.008	0.372	0.044	0.055
	Hispanic or Latino	0.017	0.501	0.126	0.105
	Median age	31.800	44.700	39.055	2.237
	Dependency Ratio	41.343	61.608	54.971	3.510
	Sex Ratio (Males per 100 Females)	90.900	109.600	98.692	3.406
Education	Pop. with Bachelor's degree or higher, age 25+	0.241	0.630	0.347	0.068
Urbanization	Urban Population (%)	0.351	1.000	0.730	0.152
Economic / Occupational	Median household income	48716.00	90203.00	69240.24	11313.50
	Full-time, year-round civilian employed population age 16+	0.348	0.484	0.405	0.029
	Workers that work from home age 16+	0.038	0.198	0.092	0.027
	Annual Change in Real GDP	0.300	9.300	5.061	2.246

	Prof. Scientific Technical Services (PSTS) Employment	0.035	0.247	0.084	0.035
Infrastructure	Pop. living where broadband fixed services of at least 25/3 Mbps are available	0.803	0.992	0.941	0.047
Social Capital	Social Capital Index	0.000	4.233	2.152	1.000
Societal Openness	Freedom Index	0.000	1.405	0.894	0.279
Affordability	Main Reason Not Online at Home: Too Expensive	0.006	0.024	0.012	0.004
	n=51 (50 states & Washington, D.C.)				

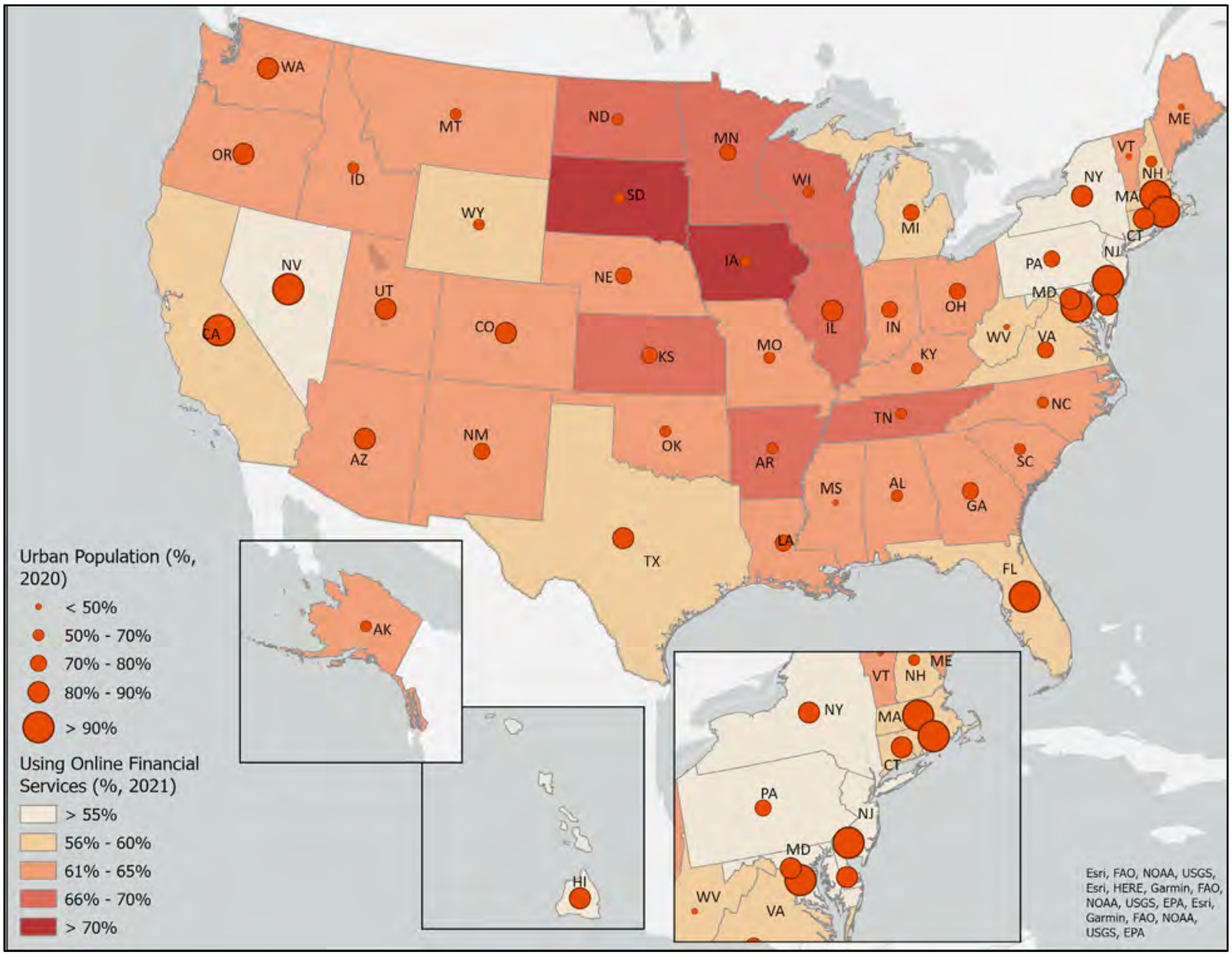
GEOGRAPHIC PATTERNS OF PURPOSEFUL INTERNET USE

All sixteen dependent indicators of online activities were mapped using a GIS. GIS mapping allows descriptive visualization of the spatial patterns of each online activity. These patterns often provide useful visual cues on the extent of agglomeration of online activities in U.S. states and reveal clusters and outliers. This mapping of the study's dependent variables is consistent with geographic mapping of broadband availability and coverage in the United States, an ongoing initiative of the FCC. In fact, broadband data collection and its mapping remain a key part of the 2021 Infrastructure Investment and Jobs Act to help identify areas lacking adequate broadband coverage, prioritize regions for infrastructure investments and provision of reduced-cost broadband services in underserved regions and impoverished communities.

Figure 3 shows the geographic distribution of purposeful use of the internet to access online financial services (online banking, investing, and bill payment) in U.S. states. The map of this online activity reveals that north-eastern states such as New York, New Jersey, Pennsylvania, and Maryland along with Nevada in the West reported the lowest usage (by no more than 55% of the state population age 15+) in November 2021. 56-60% of the population in populous states such as California, Texas, Florida, Michigan, as well as Virginia and West Virginia, and neighboring northeastern states such as Connecticut, Massachusetts, and New Hampshire engaged in this online activity during the same period. Moderate levels of engagement (61-65% of the state population age 15+) was most prevalent across U.S. states and was observed for states in the Pacific Northwest, Rocky Mountains, Southwestern US, extended parts of the deep south stretching from Louisiana in the Gulf Coast to South Carolina in the Atlantic coast. Moderate levels were also noticed for parts of the Midwest (Indiana, Ohio) and the Appalachia (Kentucky). High levels of use of online financial services (66-70% of the population age 15+) was observed in a couple of southern states (Arkansas and Tennessee), Midwest and upper Midwest (from Illinois to Minnesota), while the highest use (70-87% of the state population) was reported for Iowa in the Great Plains, South Dakota in the Prairie region, and Washington, D.C.

While Washington, D.C.'s population is 100% urban, it is evident from Figure 3 that there seems to be a reversal in terms of the urban-rural divide that is often associated with broadband and internet use in the United States. In fact, the laggards (states such as NY, NJ, CA, FL, MA, CT) have high to very high levels of urban population per capita (often exceeding 70 percent of the state population). Conversely, the leaders with the exception of Washington D.C. have much lower urban population per capita. This reversal, revealed by overlaying independent variables of interest on a dependent variable, is also noticed for other online activities and reveals an interaction between a dependent variable – in this case, online financial activities, with a potential explanatory variable, in this case, urban population per capita. This shows the importance as well as the potential of GIS mapping for this study and digital divide research in general.

Figure 3: Geographic Patterns of Internet Use for Online Financial Services, along with urban population, U.S. states, November 2021 (Data Source: NTIA, U.S. Census Bureau)



Clusters of Online Activities

K-means clustering, an unsupervised data mining technique was used to divide the sample of n = 51 U.S states plus Washington, D.C. into five (k = 5) different clusters for each category of online activities. The K-means algorithm minimizes dispersion within clusters, for example, for communication, it divides states into 5 clusters that are similar in their internet use for emailing, text and instant messaging, social networking, and interacting with household equipment using the internet. Different experiments were conducted with values of k = 4, 5, and 6 and ultimately k = 5 was selected since it produced agglomerations that were not overwhelmingly dominant and no one cluster accounted for a vast majority of U.S. states.

For the communication category, cluster 1, the lowest use cluster was comprised of six states – Delaware, New York, and Pennsylvania in the northeast, the sole southern state of Mississippi, West Virginia in the Appalachian region, and Hawaii in the west. Internet use for communication was lowest in the states in this cluster as evident from the cluster center values for the four indicators of e-communication in Table 2. Cluster 2 has low use of the internet for communication and was comprised of 15 states. Among these 15 states were California and Nevada in the west, Texas, Alabama, and Florida, Midwestern states in the Great Lakes region such as Michigan and Ohio, Kentucky and Virginia in the Appalachia, and several northeastern including Connecticut, Maryland, Massachusetts, New Hampshire, and Rhode Island.

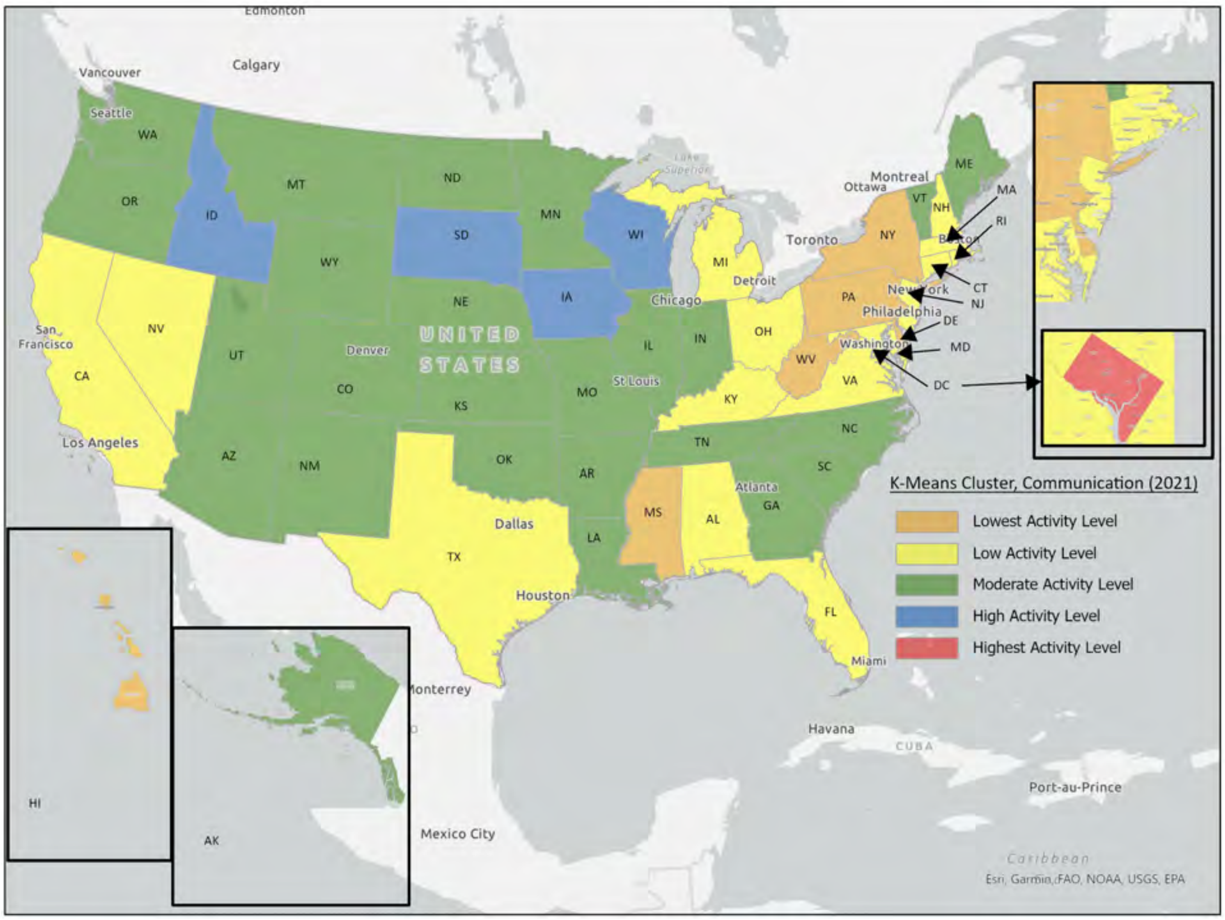
Table 2: K-means Clusters of Online Activities for Communication, with Cluster Centers

	Lowest Activity Level	Low	Moderate	High	Highest Activity Level	
Cluster	1	2	3	4	5	Ratio = MAX/MIN
Email	0.679	0.724	0.777	0.842	0.997	1.47
Text msg/Instant msg	0.700	0.731	0.786	0.856	0.992	1.42
Social Networking	0.547	0.577	0.634	0.692	0.814	1.49
Interact with household Equipment	0.130	0.181	0.179	0.175	0.254	1.96
n	6	15	25	4	1	
	Delaware Hawaii Mississippi New York Pennsylvania West Virginia	Alabama California Connecticut Florida Kentucky Maryland Massachusetts Michigan Nevada New Hampshire New Jersey Ohio Rhode Island Texas Virginia	Alaska Arizona Arkansas Colorado Georgia Illinois Indiana Kansas Louisiana Maine Minnesota Missouri Montana Nebraska New Mexico North Carolina North Dakota Oklahoma Oregon South Carolina Tennessee Utah Vermont Washington Wyoming	Idaho Iowa South Dakota Wisconsin	District of Columbia	

Cluster 3 was the largest cluster with 25 states spanning the entire nation. These states had moderate levels of e-communication related online activities. The states spanned Washington and Oregon in the Pacific Northwest, Montana to Minnesota in the Prairies, large swathes of the south From Oklahoma to South Carolina, Illinois, Indiana in the Midwest and Vermont and Maine in the northeast. Clusters 4 and 5 had high and highest levels of online activities for communication. Cluster 4 was comprised of Idaho, South Dakota, Iowa, and Wisconsin. Cluster

5 had only one member, Washington, D.C. which had the highest level of e-communication using the internet. In other words, Washington D.C. may be viewed as a high outlier. The clusters were mapped using a GIS and the geographic arrangement of clusters of e-communication related online activities shows that oftentimes, states belonging to the same cluster are contiguous with some exceptions (Figure 4). This indicates spatial spillover effects and is consistent with Tobler's Law, which states ""Everything is related to everything else, but near things are more related than distant things (Tobler, 1970)."

Figure 4: K-means clusters of Internet use for Communication, U.S. States, Nov 2021



The spatial agglomeration of states within clusters of e-communication implies possible presence of spatial bias, which may unduly influence the analysis of socioeconomic determinants of online activities. This has methodological implications and as discussed in the previous section calls for the analysis of spatial autocorrelation to account for spatial bias of the sixteen dependent variables (Longley et al., 2015).

The clusters of online activities pertaining to communication were characterized in terms of their demographic, social, economic, and other independent attributes. Cluster characteristics in Table 3 show that states in clusters 1 and 2 (low to lowest levels of online activities for communication) compared favorably with states in clusters 3 and 4 (high to highest levels) in terms of educational attainment (population with Bachelors degree or higher) but were slightly older, with a difference of almost 2 years in median age. Interestingly, these states tended to have higher proportions of minority populations including Blacks or African Americans, Asians, and Hispanics. Interestingly, the 21 states in clusters 1 and 2 had comparable or higher urban population per capita compared to the 29 states in clusters 3 and 4. This is contrary to well-documented discrepancy in urbanization in states that lag ICT adoption and use. In terms of economic and occupational outlook, states in clusters 1 and 2 had comparable civilian employed population age 16+ compared to states in clusters 3 and 4. The same is largely true for remote workers per capita. Interestingly however, these clusters reported slightly higher proportion of workers in professional, scientific, and technical services occupations.

Overall, k-means cluster characteristics reveal important differences between clusters in extent of urbanization and minority population as well as social interconnectedness (measured by social capital index) and societal openness (measured by the Freedom Index). The clusters are largely comparable in terms of age, educational attainment, and economic and occupational attributes including proportion of remote workers, PSTS employment, and median household income. The clusters are also comparable in terms of infrastructure measured by broadband coverage of population. Implications of these differences and similarities are discussed later in this paper.

Table 3: Characteristics of 5 Clusters of Online Activities for Communication						
		Lowest Activity Level	Low	Moderate	High	Highest Activity Level
Category of Ind. variable	Variable	1	2	3	4	5
Demographic	Black or African American	14.38%	12.29%	8.87%	3.13%	43.18%
	Asian	9.21%	5.42%	2.96%	2.08%	4.13%
	Hispanic or Latino (of any race)	8.99%	16.68%	11.83%	7.90%	11.48%
	Median age (years)	40.65	39.76	38.528	38.375	34.8
	Dependency Ratio	57.3465	53.4681	55.3215	58.2610	41.3435
	Sex Ratio (Males per 100 Females)	97.0333	97.5067	99.7440	101.0000	90.9000
Education	Population with Bachelor's degree or higher, age 25+	32.38%	35.99%	33.83%	31.37%	63.05%
Urbanization	Urban Population (%)	70.58%	80.78%	69.20%	64.17%	100.00%
Economic / Occupational	Median household income ($)	66530.5	74186.87	66553.44	66335.5	90088.0
	Full-time, year-round civilian employed population 16+	37.65%	40.55%	40.54%	42.77%	48.42%
	Labor force participation rate	59.75%	63.51%	63.43%	65.78%	70.80%
	Workers that work from home	7.60%	9.43%	9.08%	8.38%	19.80%
	Annual Change in Real GDP	4.15	6.34	4.544	5.2	3.7
	PSTS Employment	7.08%	9.59%	7.73%	5.87%	24.68%

Infrastructure	Households with a broadband subscription (%)	84.92%	87.57%	85.98%	86.13%	87.20%
	Pop. living where broadband fixed services of at least 25/3 Mbps are available (%)	92.13%	96.36%	93.01%	94.88%	98.00%
Social Capital	Social Capital Index	1.569	1.878	2.310	3.248	1.449
Societal Openness	Freedom Index	0.571	0.917	0.933	1.069	0.786
Affordability	Main Reason Not Online at Home: Too Expensive	1.41%	1.31%	1.22%	0.81%	1.71%

K-means clusters of online activities categorized as entertainment, education and work, and finance and commerce were also determined. While the cluster composition for online activities spanning education and work and finance and commerce resemble that of communication, there are differences for online activities representing entertainment. For example, the lowest cluster (cluster 1) in terms of entertainment activities has 13 states followed by another 21 states in the low cluster (cluster 2). In other words, these two clusters account for over two-thirds of U.S. states. Clusters 3, 4, and 5 with moderate, high, and highest levels of entertainment online activities had a total of 16 states plus Washington, D.C., the sole member and outlier in cluster 5. Cluster composition and cluster centers are reported in Table 4 for entertainment online activities. Characteristics of clusters for entertainment online activities as well as work and education and finance and commerce are however mostly comparable to those of clusters for communication related online activities.

Table 4: K-means Clusters of Online Activities for Entertainment, with Cluster Centers

Cluster	1	2	3	4	5	Ratio = MAX/MIN
Publishing or Uploading Blog Posts, Videos, or Other Original Content	0.126	0.129	0.123	0.154	0.210	1.66
Watching Videos Online	0.516	0.568	0.622	0.661	0.865	1.68
Streaming Downloading Music, Video, Podcasts	0.435	0.488	0.524	0.590	0.795	1.83
n	13	21	11	5	1	
	Alabama Delaware Florida Georgia Hawaii	California Connecticut Indiana Iowa Kansas	Alaska Arizona Arkansas Illinois Minnesota	Colorado Idaho Oregon South Dakota	District of Columbia	

	Kentucky Mississippi New Jersey New York North Carolina Pennsylvania Rhode Island West Virginia	Louisiana Maine Maryland Massachusetts Michigan Missouri Nevada New Hampshire North Dakota Ohio Oklahoma South Carolina Tennessee Texas Vermont Wyoming	Montana Nebraska New Mexico Virginia Washington Wisconsin	Utah		

For the sake of brevity, additional discussion of clusters and their characteristics for other categories of online activities are excluded from this paper.

Spatial autocorrelation analysis did not reveal any major issues. Only two online activities – Request Services Provided by Other People via the Internet and Offer Services for Sale via the Internet, were found to be significantly clustered with Moran's I values of .167 and .131, significant at the .001 and .01 levels (bottom of Table 6). Hence it is essential that OLS regression models for these two variable account for the presence of spatial bias.

SOCIOECONOMIC ANALYSIS OF PURPOSEFUL INTERNET USE

Ordinary Least Squares (OLS) regressions were conducted in a stepwise manner to determine associations of independent variables with the sixteen dependent indicators of online activities. With a sample size of n = 51 (50 states plus Washington, D.C.), care was taken to enter no more than 5 independent variables into each OLS regression.

Regression findings (shown in Tables 5 & 6) reveal that the dominant correlates of online activities in U.S. states are percent of remote workers and median age. Remote workers were positively associated with 12 of the 16 dependent variables, across all four categories – communication, entertainment, work and education, and financial and commercial online activities. In other words, as more people worked from home during the pandemic, they increasingly relied upon the internet for e-communication, e-entertainment, teleworking and e-education, as well as e-commerce. While intuitive, this finding is novel since this is the first time the positive influence of remote work with purposeful internet use has been documented for the United States. The high standardized beta coefficients, all significant at the .001 level, reinforce the strong positive relationship between remote work and the extent of purposeful internet use in U.S. states. Similar to remote workers, median age was also associated (inversely) with 12 of the 16 dependent variables across all four categories. The inverse association in each case shows that as age increases, the likelihood of engaging in online activities decreases. The inverse association of age with the adoption and use of information and communication technologies (ICTs) such as broadband and the internet has been well-documented in the digital divide literature. In the United States, age has been documented to be a barrier to computer and internet use at home (NTIA, 2011) and digital knowledge and skills have been confirmed to vary substantially by age (Vogels and Anderson, 2019). It has been reasoned in a recent study (Mubarak and Suomi, 2022) that due to a lack of digital skills, more advanced age cohorts are unable to connect with peers leading to social exclusion, thereby increasing gaps in purposeful internet use. In fact, the pandemic is likely to have exacerbated social exclusion thus widening gaps in internet use across the spectrum of online activities among older age groups.

Proportion of urban population is found to be inversely associated with using email, text messaging and instant messaging, watching videos online, and streaming or downloading music, radio, podcasts at the .01 level or lower. However, urban population is positively associated with interacting with household equipment using the internet. The inverse association of urban population with four dependent variables is certainly noteworthy and contrary to findings from a prior study that found urban to be positively associated with desktop/laptop use, internet access, and broadband use in the household in U.S. states (Pick, Sarkar, and Johnson, 2015).

Table 5: Standardized Regression Results – Communication and Entertainment Online Activities									
		Communication / IoT					*Entertainment*		
		Use Email	Text Msg or Instant Msg	Online Social Networks	Interact with HH Equipment via Internet		Watching Videos Online	Stream or Download Music, Radio, Podcasts	Publish or Upload Blog Posts, Videos, or Other Original Content
Demographic	Black or African American								
	Asian			-0.472***					-0.240*
	Hispanic or Latino (of any race)								
	Median age	-0.349**	-0.554***	-0.409***			-0.507***	-0.547***	-0.390***
	Dependency Ratio								
	Sex Ratio (Males per 100 Females)								
Education	Population with Bachelor's degree or higher, age 25+				-0.727***				
Urbanization	Urban Population (%)	-0.481***	-0.558***		0.351***		-0.383**	-0.357***	
Economic / Occupational	Full-time, year-round civilian employed population 16+	0.221			0.552***				
	Workers that work from home	0.605***	0.554***		0.812***		0.717***	0.746***	0.526***
	Annual Change in Real GDP								
	PSTS Employment								
Infrastructure	Pop. living where broadband fixed services of at least 25/3 Mbps are available (%)								
Social Capital	Social Capital Index								-0.354**
Societal Openness	Freedom Index								
Affordability	Main Reason Not Online at Home: Too Expensive								
	Adjusted R^2	0.505***	.441***	0.365***	.636***		.555***	.640***	.445***
	VIF	1.751	1.477	1.000	4.956		1.477	1.477	1.109
	Spatial Autocorrelation								

	Moran's I - Dep Var	-0.086	-0.069	-0.092	0.074		-0.023	-0.087	-0.02
	Moran's I - Residual	-0.078	-0.094	-0.214***	-0.020		-0.046	-0.142*	-0.071
	** p<.05, **p<.01, ***p<.001*								

Table 6: Standardized Regression Results – Work & Education and Financial and Commercial Activities

		Work/Education					*Financial/Commercial*				
		Remote Work	Online Job Search	Take Classes or Participate in Job Training Online	Participate in Online Video or Voice Calls or Conferenc es		Use Online Financial Services (Banking, Investing, Paying Bills, etc.)	Shop, Make Travel Reservations or Use Other Consumer Services Online	Sell Goods Online via the Internet	Request Services Provided by Other People via the Internet	Offer Services for Sale via the Internet
Demographic	Black or African American			-0.260***							
	Asian	-0.198*						-0.244*	0.350**		
	Hispanic or Latino (of any race)							-0.229*	-0.282*		
	Median age	-0.242	-0.288**	-0.461***			-0.227*			-0.284***	-0.255*
	Dependency Ratio										
	Sex Ratio (Males per 100 Females)							0.435***			
Education	Population with Bachelor's degree or higher, age 25+	0.839***									
Urbanization	Urban Population (%)			-0.127							
Economic / Occupational	Full-time, year-round civilian employed population 16+		0.237*								
	Workers that work from home		0.501***	0.885***	0.899***		1.150***	0.777***		0.563***	
	Annual Change in Real GDP				-0.286***		-0.369***				
	PSTS Employment						-0.495**			0.280*	0.520***

Infrastructure	Pop. living where broadband fixed services of at least 25/3 Mbps are available (%)			-0.288**							
Social Capital	Social Capital Index								0.576***		
Societal Openness	Freedom Index						0.203*				
Affordability	Main Reason Not Online at Home: Too Expensive										
	Adjusted R^2	.730***	.540***	.765***	.704***		.563***	.609***	.328***	.788***	.346**
	VIF	1.054	1.431	2.193	1.130		4.728	1.121	1.224	4.310	1.020
	Spatial Autocorrelation										
	Moran's I - Dep Var	0.170***	0.027	-0.027	0.072		0.131	0.038	0.058	0.167***	0.131**
	Moran's I - Residual	0.041	-0.063	-0.088	-0.102		-0.055	-0.062	-0.097	-0.024	-0.082
	** p<.05, **p<.01, ***p<.001*										

We reason that due to population migration that took place during the pandemic from densely populated, large urban metropolitan areas to less densely populated micropolitan or rural areas, an increase in proportion of urban population may have contributed to reducing the extent to which people in urban areas used the internet for online activities. This was evident from Figure 3 which showed that higher levels of using online financial services for banking, investing and bill payment often happened in states with lower proportions of urban population. The exception is the dependent variable interacting with household equipment using the internet, a more contemporary online activity that requires higher levels of digital and internet skills.

Among the race/ethnicity variables, Asian population is inversely associated with four dependent variables – using online social networks, publishing or uploading content using the internet such as blogs and videos, working remotely using the internet, and e-commerce. However, Asian population is found to be positively associated with requesting services provided by other people over the internet. For use of online social networks and publishing or uploading content using the internet, Asians have trailed the other race/ethnic groups including White, non-Hispanic, African Americans, and Hispanics. For the other two dependent variables, Asians have led the other race/ethnic groups and hence those regressions findings merit additional research. For selling goods via the internet, Asian Americans were the leading race/ethnic group and as such, the positive association makes sense. Among the other independent variables, only employment in professional, scientific, and technical services employment is associated with financial and commercial activities, for 3 of the 4 indicators. Otherwise, OLS regressions do not show any meaningful and consistent association of infrastructure, social capital, societal openness, and affordability with a broad spectrum of the dependent variables.

This set of OLS regressions are found to explain 44.1%-78.8% variation of 13 out of 16 dependent variables showing the robustness of the associations proposed in the conceptual model of this study. The Variance Inflation Factor (VIF) in each regression does not exceed a cutoff of 5.0 showing that multicollinearity is not a concern.

For the two variables that were clustered (Moran's I was statistically significant for Request Services Provided by Other People via the Internet Offer Services for Sale via the Internet), we find that the regression residuals are geographically randomly distributed (Moran's I values of -.024 and -.082, neither significant, reported in bottom of Table 6). We can conclude that both regression models were able to account for the presence of spatial bias indicating the robustness of the posited associations.

Recognizing the dominance of proportion of remote workers as an independent correlate, a second set of OLS regressions were conducted excluding this independent variable to determine any other independent variable that may have significantly influenced purposeful internet use in the United States during the pandemic. This set of OLS regressions (results in Tables 7 & 8) reinforced the dominance of median age which continued to be inversely associated with 11 of the 16 dependent variables. Other important correlates are educational attainment, specifically proportion of population age 25+ with a Bachelors degree or higher. This variable was found to be positively associated with 7 out of the 16 online activities spanning all four categories. Higher levels of educational attainment are often associated with higher levels of digital skills and perceived use of the internet. In fact, the positive association of educational attainment has been extensively documented in the digital divide literature (Pick and Sarkar, 2015, van Dijk 2020, and Warschauer and Matuchniak 2010).

Table 7: Standardized Regression Results *without* Workers that Work from Home – Communication and Entertainment Online Activities

		Communication / IoT					*Entertainment*		
		Use Email	Text Msg or Instant Msg	Online Social Networks	Interact with HH Equipment via Internet		Watching Videos Online	Stream or Download Music, Radio, Podcasts	Publish or Upload Blog Posts, Videos, or Other Original Content
Demographic	Black or African American								
	Asian		-0.248	-0.472***	-0.288**				
	Hispanic or Latino (of any race)								
	Median age		-0.516***	-0.409***			-0.423***	-0.400***	-0.513***
	Dependency Ratio								
	Sex Ratio (Males per 100 Females)								-0.301*
Education	Population with Bachelor's degree or higher, age 25+		0.410**				0.360**		0.292*

Urbanization	Urban Population (%)		-0.349*		0.571***				
Economic / Occupational	Full-time, year-round civilian employed population 16+	.532***			0.435***				
	~~Workers that work from home~~								
	Annual Change in Real GDP								
	PSTS Employment							0.503***	
Infrastructure	Pop. living where broadband fixed services of at least 25/3 Mbps are available (%)								
Social Capital	Social Capital Index						0.262*	0.282**	
Societal Openness	Freedom Index							0.207*	
Affordability	Main Reason Not Online at Home: Too Expensive								
	Adjusted R^2	.268***	.375***	.365***	.566***		.422***	.545***	.346***
	VIF	1.000	1.701	1.000	1.365		1.111	1.079	1.223
	** p<.05, **p<.01, ***p<.001*								

Table 8: Standardized Regression Results *without* Workers that Work from Home – Work & Education and Financial and Commercial Activities

		Work/Education					*Financial/Commercial*				
		Remote Work	Online Job Search	Take Classes or Participate in Job Training Online	Participate in Online Video or Voice Calls or Conferences		Use Online Financial Services (Banking, Investing, Paying Bills, etc.)	Shop, Make Travel Reservations, or Use Other Consumer Services Online	Sell Goods Online via the Internet	Request Services Provided by Other People via the Internet	Offer Services for Sale via the Internet
Demographic	Black or African American			-0.355***							
	Asian	-0.198**		-0.197*					-0.282*		
	Hispanic or Latino (of any race)								0.350**		
	Median age	-0.242**	-0.246*	-0.466***	-0.309***					-0.289***	-0.255*
	Dependency Ratio										
	Sex Ratio (Males per 100 Females)										
Education	Population with Bachelor's degree or higher, age 25+	0.839***		0.832***	0.867***					0.344*	

Urbanization	Urban Population (%)				-0.208*						
Economic / Occupational	Full-time, year-round civilian employed population 16+		0.311**				.511***				
	~~Workers that work from home~~										
	Annual Change in Real GDP										
	PSTS Employment		0.408***					0.416***		0.470**	0.520***
Infrastructure	Pop. living where broadband fixed services of at least 25/3 Mbps are available (%)			-0.292**							
Social Capital	Social Capital Index							0.511***	0.576***		
Societal Opennes	Freedom Index										
Affordability	Main Reason Not Online at Home: Too Expensive										
	Adjusted R^2	.730***	.477***	.691***	.690***		.246***	.420***	.328***	.737***	.346***
	VIF	1.054	1.339	1.602	1.414		1.000	1.000	1.224	4.514	1.020
	** p<.05, **p<.01, ***p<.001*										

Following educational attainment, professional, scientific, and technical services (PSTS) employment is found to be positively associated with 5 of the 16 dependent variables – mostly in the financial and commercial category of online activities. We reason that Americans employed in occupations such as computer and mathematical science, legal, business and financial services, life, physical and social sciences, and management are likely to have higher needs to use digital technologies including the internet. In addition, teleworking has continued to grow for this group of employed Americans (NTIA, 2019) explaining the positive association of PSTS employment with online activities. This set of regressions also reveals that social capital is positively associated with watching videos online, streaming or downloading music, radio, and podcasts, as well as online shopping, making travel reservations, and selling goods online. The positive influence of social capital signifies the role of social interconnectedness as well as the transfer of digital skills due to communal connections for online entertainment and financial and commercial activities. Like the previous set, among race/ethnic variable, Asian population remains inversely associated with 6 of the 16 dependent variables. Overall, when remote workers is excluded as an independent variable, it is replaced by educational attainment (Bachelors degree or higher) and PSTS employment, as an economic indicator. The influence of social capital on online activities that are related to entertainment and commerce also comes to the forefront.

Overall, OLS regressions reveal the importance of remote working on online activities in U.S. states, a novel, yet intuitive finding. As more Americans have worked from home since 2019 and the pandemic fueled greater remote working for certain occupations, the appetite to use the internet in a variety of ways to communicate with people and/or household equipment, seek education, entertainment, and access to online financial services has certainly expanded, as reflected in this finding. This has implications for infrastructure development and deployment as well as for the need to advance digital knowledge and skills. Age related discrepancies continue to plague the U.S. digital divide, as evident from the strong inverse association of median age with online activities in both sets of regressions. Among other variables, Asian population, Bachelors education, PSTS employment, and social capital show a socio-demographic as well economic influence on the digital divide in the United States. The implications of these findings are discussed next.

IMPLICATIONS OF FINDINGS

The findings have some implications that are commonly known and other new ones. The implications of the correlates of digital purposeful uses will be discussed first, following by the implications of spatial patterns and agglomerations. The strong association of remote work with variables across the four dependent categories has many implications that are known through the almost pervasive experience of the pandemic and beginning to be known through published studies that are appearing more and more.

The pervasiveness of impacts has been confirmed reported from interviews and surveys of people sheltering in three advanced nations (US, Germany, Switzerland) during the pandemic (Hargittai, 2022). For instance, social media was particularly strongly associated with many aspects of peoples' lives, including outcomes on health, professions, investments, and political participation (Hargittai, 2022). What is less know is the directionality of effect, i.e did social media use cause impacts and often improvements in life experiences, or was the reverse true, that during the pandemic people's health, financial, and other needs caused more social media use. Our findings reinforce quantitatively the association of remote work with expanded use of social media, online entertainment, online job-related activities, and online finance and commerce, while we did not confirm the causality directionality, which might well have been bidirectional. The implication is that remote work status has greatly altered peoples' ways of information gathering, communicating, investing, doing work, and participating politically. This is important not only historically during the pandemic, but for patterns that are continuing perhaps in moderated form post-pandemic. This implication points to the need for future studies to assess the level of persistence of the covid online lifestyle.

We explain that the positive association of urban location with digitalization of household equipment was due to the needs of people in isolation and sheltering to focus more on upgrading household equipment with a product mix more likely to have capabilities for internet communications, such as contemporary security systems, utility controls, landscaping online connections, and kitchen devices, which would be more associated with retail and service entities in urban areas. On the other hand, urban location was inversely associated with other online activities of e-mail use, texting and instant messaging, online video watching, and streaming music and podcasts. We posit that an implication is that rural areas have been able to somewhat “catch up” with urban areas in the online activities in our study. The research literature on digital divides has predominantly found urban location to be a positive influence on digital uses (Pick and Sarkar, 2015, Lythreatis et al., 2022), and rural areas may have been challenged by less opportunity for learning and training of digital skills (Hodge et al., 2017). However, for these findings on inverse associations, an implication may be that rural residents, by the necessity of the covid pandemic, learned much more about online activities. Another implication is that migration away from urban centers stimulated by the pandemic, may continue in the future, with technically skilled migrants adding to the digital capabilities of rural communities.

The consistently inverse association of median age with purposeful online uses corresponds to a well-known effect of age that is pervasive in the digital divide literature. The implication is that older people can benefit by training in digital skills and in knowledge of online resources that can benefit them. In the post-pandemic era, the training may be easier through online means, and the motivations of older people may be higher given the digital transformation of society that is an overall outcome of the pandemic. One other point is that although the age-related digital divide remains from the findings of this study, the overall digital capabilities throughout the age groups has risen during covid, implying more digital opportunities for older people.

Regarding the spatial implications of this research, the shift in rural-urban associations implies that for some online activities, rural areas of the US during the pandemic had among the highest levels of online purposeful uses. This was illustrated in Figure 3, in which using online financial services was very high in the Prairie States of South Dakota and Iowa, surprisingly higher than traditional urban leaders such as Massachusetts, New York, and New Jersey, with the caveat that Washington DC remained very high. The lowest states shown in Figure 3 include Nevada, Pennsylvania, and New York. These geographic reversals were common for a number of other online purposeful activities. The implications of this reversal for businesses is that online marketing of products may need to shift considerable in focus, compared to pre-pandemic. There are also e-government implications at the state and federal levels, so that it can be expected that many online government services can be provided with greater use and acceptance in rural states. For the future, studies need to be done post-pandemic, to determine if this new balance of geographic associations will persist or return to pre-pandemic traditional geographic patterns favoring urban and metropolitan states and counties in online purposeful uses.

CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

This study analyzes geographic patterns and socioeconomic influences on online activities that constitute purposeful internet use in U.S. states during the COVID-19 pandemic (November, 2021). An array of sixteen online activities that span e-communication, e-entertainment, e-education, teleworking, and e-commerce provide a comprehensive overview of the digital lives of Americans and their purposeful use of the internet during the pandemic. One of the notable features of this study is the GIS mapping of online activities which provides important visual cues about fluctuations in levels of online activities in U.S. states.

Mapping has been a policy priority for federal agencies such as the FCC and NTIA and this study uses mapping to analyze state-level patterns of online activities, disparities between states, as well as clusters. Map overlays enable descriptive analysis of interactions between the extent of online activities and factors such as race/ethnicity, urbanization, income and occupation, and other variables that are traditionally key in analysis of digital divides. Clusters of U.S. states are identified for different categories of online activities.

It is found that states that excel in the use of the internet for online communication and IoT, entertainment and education access, as well as e-commerce are often less urban than their counterparts that lag in the same online activities. This is a surprising finding since the urban-rural digital divide has previously indicated that less urban and more rural areas tend to lag in their adoption and use of the internet. Regression findings reveal that remote workers are highly positively associated with the extent of online activities, while median age is inversely associated. We reason that as remote work expanded in many occupations due to the COVID-19 pandemic, the spectrum of internet use also expanded and grew for all categories of purposeful internet use. Along with teleworking, internet users accessed entertainment and education using the internet, conducted e-commerce, and of course used the internet for traditional communication with other people as well as to interact with household devices. Urbanization was found to be inversely associated with a smaller array of online activities, while educational attainment was positively associated. For a small selection of online activities, employment in professional, scientific, and technical services was found to be a positive explanatory variable.

The study is limited for its analysis of online activities at the state level. In other words, within-state fluctuations are obscured as a result. The aggregation of NTIA data at a smaller unit of U.S. Census geography, for example, counties, may yield newer and more nuanced insights about the similarities and disparities in online activities. It is also likely that due to the modifiable areal unit problem (MAUP), one of the pitfalls of geographic data (O'Sullivan and Unwin, 2010), the findings of this study regarding demographic and socioeconomic associations may be affected. Hence a separate and standalone county-level study is proposed as a future research direction to analyze geographic patterns of online activities and their demographic and socioeconomic influences.

Finally, while this study has a broad array of variables that constitute purposeful internet use during the COVID-19 pandemic, it does not include any online activity related to electronic health and e-government at the present time. For example, during the pandemic, finding health- and insurance-related information, communicating with doctors and nurses online, and using electronic health monitoring devices that are connected to the internet became quite prevalent. Accessing government services using the internet started growing before the pandemic and was certainly accelerated. As data on such variables continues to be published by agencies such as the NTIA, their inclusion will expand the scope of the current study and provide a more holistic understanding of purposeful internet use in American households. This is also outlined as a direction for future research.

Last but not the least, to gain a fuller understanding of the effects on the pandemic on the digital lives of Americans, it is essential to compare the findings of this study based on 2021 data on online activities with pre-pandemic timeframes (for example, 2019 and 2017 data on online activities, documented by the NTIA). A longitudinal comparative study is also outlined for future research.

REFERENCES

Agarwal, R., Animesh, A., & Prasad, K. (2009). Research note—Social interactions and the "digital divide": Explaining variations in internet use. Information Systems Research, 20(2), 277-294.

Azari, R., & Pick, J. B. (2005). Technology and society: socioeconomic influences on technological sectors for United States counties. *International Journal of Information Management*, *25*(1), 21-37.

Chen, W. 2013. The implications of social capital for the digital divides in America. *The Information Society* 29:13-25.

Hargittai, E. 2022. *Connected in Isolation: Digital Privilege in Unsettled Times*. Cambridge, MA: MIT Press.

Hodge, H, Carson, D., Carson, D., Newman, I, and Garrett, J. 2017. Using internet technologies in rural coommunities to access services: the views of older people and service providers. Journal of Rural Studies 54: 469-478.

Longley, P.A, Goodchild, M.F., Maguire, D.J, and Rhind, D.W. (2015). *Geographic Information Systems and Science*, 4th ed., New York, NY: John Wiley and Sons.

Lythreatis, S., Singh, S.K., and El-Kassar, A.-N. 2022. The digital divide: A review and future research agenda. *Technological Forecasting and Social Change.* 175:1-11.

Moran, P. A. (1950). A test for the serial independence of residuals. *Biometrika*, *37*(1/2), 178-181.

Mossberger, K., Tolbert, C.J., Stansbury, M., 2003. *Virtual Inequality: beyond the digital divide*. Georgetown University Press.

Mubarak, F., & Suomi, R. (2022). Elderly forgotten? Digital exclusion in the information age and the rising grey digital divide. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 59, 00469580221096272.

Myers, R.H. 1990. *Classical and Modern Regression Application*. 2nd edition. Duxbury press. CA.

NTIA. (2022). Digital Nation Data Explorer. National Telecommunications and Information Administration, Washington, D.C., https://ntia.gov/other-publication/2022/digital-nation-data-explorer#sel=internetUser&disp=map.

NTIA. (2019). Nearly a Third of American Employees Worked Remotely in 2019, NTIA Data Show, National Telecommunications and Information Administration, Washington, D.C., Retrieved from https://www.ntia.gov/blog/2020/nearly-third-american-employees-worked-remotely-2019-ntia-data-show.

NTIA. (2011). Exploring the Digital Nation: Computer and Internet Use at Home. National Telecommunications and Information Administration, Washington, D.C. Retrieved from https://www.ntia.gov/sites/default/files/publications/exploring_the_digital_nation_computer_and_internet_use_at_home_11092011_0.pdf.

NTIA. (2013). Exploring the Digital Nation: America's Emerging Online Experience. National Telecommunications and Information Administration, Washington, D.C. Retrieved from https://www.ntia.doc.gov/files/ntia/publications/exploring_the_digital_nation_-_americas_emerging_online_experience.pdf.

OECD (2001), "Understanding the Digital Divide", OECD Digital Economy Papers, No. 49, OECD Publishing, Paris, https://doi.org/10.1787/236405667766.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and planning A*, *16*(1), 17-31.

O'Sullivan, D., & Unwin, D. J. (2010). *Geographic Information Analysis*. John Wiley & Sons, Hoboken, NJ.

Pick, J. B., & Sarkar, A. (2015). *The global digital divides: Explaining change*. Berlin: Springer.

Pick, J., Sarkar, A., & Rosales, J. (2019). Social media use in American counties: geography and determinants. *ISPRS International Journal of Geo-Information*, 8(9), 424.

Pick, J. B., Sarkar, A., & Johnson, J. (2015). United States digital divide: State level analysis of spatial clustering and multivariate determinants of ICT utilization. *Socio-Economic Planning Sciences*, 49, 16-32.

Rogers, E. M. (2010). *Diffusion of Innovations*. Simon and Schuster. New York, NY.

Ruger, W.P., and Sorens, J. 2021. An index of personal and economic freedom. Cato Institute, Washington, D.C. Retrieved from https://cdn.freedominthe50states.org/download/2021/print-edition-2021.pdf.

Sarkar, A. Pick, J.B. and Bhat, G. (2022). Purposeful Mobile Internet Adoption and Use in the United States: A Geodemographic and Socioeconomic Analysis. In Proceedings of the Telecommunications Policy Research Conference (TPRC), Washington, D.C., available at http://dx.doi.org/10.2139/ssrn.4178779.

Sarkar, A., Pick, J. B., & Rosales, J. (2023). Multivariate and geospatial analysis of technology utilization in US counties. *Telecommunications Policy*, 47(7). https://doi.org/10.1016/j.telpol.2023.102572

Scarpi, D. (2012). Work and fun on the internet: the effects of utilitarianism and hedonism online. *Journal of Interactive Marketing*, *26*(1), 53-67.

Stebbins, R.A. 2001. Exploratory Research in the Social Sciences. Qualitative research methods series 48. Thousand Oaks, CA: SAGE Publishing.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1), 234-240.

U.S. Congress, 2021. Infrastructure Investment and Jobs Act. Public Law 117–58, 117th Congress, Washington, D.C.

U.S. Senate, 2018. The Geography of Social Capital in America SCP REPORT NO. 1-18, Joint Economic Committee, Washington, D.C. Retrieved from https://www.jec.senate.gov/public/_cache/files/e86f09f7-522a-469a-aa89-1e6d7c75628c/1-18-geography-of-social-capital.pdf.

Van Deursen, A. J., Helsper, E. J., & Eynon, R. (2016). Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society*, 19(6), 804-823.

Van Deursen, A., Van Dijk, J., 2011. Internet skills and the digital divide. *New Media Soc.*, 13 (6), 893–911.

Van Dijk, J.A.G.M. 2005. *The Deepening Divide: Inequality in the Information Society*. Thousand Oaks, CA: SAGE Publications.

Van Dijk, J. A.G.M. 2020. *The Digital Divide*. Cambridge, UK: Polity Press.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425-478.

Vicente, M. R., & López, A. J. (2011). Assessing the regional digital divide across the European Union-27. *Telecommunications Policy*, *35*(3), 220-237.

Vogels, E.A., and Anderson, M. 2019. Americans and Digital Knowledge. Pew Research Center. Retrieved from https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/.

Warf, B. (2012). Global geographies of the internet. Springer Science & Business Media.
Warschauer, M., and Matuchniak, T. (2010). New technology and digital worlds: analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, 34(1), 179-225.

Qualified Charitable Distributions to Split-Interest Entities

Sheldon R. Smith
Utah Valley University
smithsh@uvu.edu

ABSTRACT

The SECURE 2.0 Act of 2022, part of the Consolidated Appropriations Act of 2023, made many changes to the tax law to promote employer retirement plans and encourage retirement savings. Included in this act were some provisions that affect qualified charitable distributions (QCDs). Individuals can now make a one-time QCD to a split-interest entity. This paper describes and discusses this new opportunity.

KEYWORDS: Qualified Charitable Distributions, Split-Interest Entities, SECURE 2.0 Act

INTRODUCTION

The SECURE 2.0 Act of 2022, part of the Consolidated Appropriations Act of 2023 (P.L. 117-328), became law at the end of 2022. It made many changes to the tax law to promote employer retirement plans and encourage retirement savings. Some of the provisions included in this act relate to qualified charitable distributions (QCDs), the ability for taxpayers to make nontaxable distributions from a traditional individual retirement arrangement (IRA) to a qualifying charity. The focus of this paper will be the new provision allowing a one-time QCD to certain split-interest entities.

An example of a split-interest entity would be a trust created by a donor with charitable intentions. The donor transfers assets into the trust, which is set up to provide income to the donor for a period of years, perhaps the donor's remaining life, with the remaining value of the trust at the termination of the trust going to a designated charity. In this case, the interests in the trust are split between the lead beneficiary, the donor in this case, and the remainder beneficiary, the charity that receives the interests that remain when the trust terminates.

The Financial Accounting Standards Board (FASB) Accounting Standards Codification (ASC) defines a split-interest entity as "An agreement in which a donor enters into a trust or other arrangement under which a not-for-profit entity (NFP) receives benefits that are shared with other beneficiaries," further stating that such an agreement typically has the following two components: a lead interest and a remainder interest (ASC Master Glossary). The lead interest has the right to receive benefits from the assets during the term of the agreement, and the remainder interest has the right to receive some or all of the assets at the termination of the agreement (ASC 958-30-05-5).

The ASC further mentions the following forms that split-interest agreements can take: charitable lead annuity trust, charitable lead unitrust, charitable remainder annuity trust, charitable remainder unitrust, charitable gift annuities, and pooled income funds (ASC 958-30-05-6). Not all of these are eligible for the new QCD provision, but those that are will be discussed in more detail later. The focus of ASC 958-30 is on how not-for-profit entities account for split-interest agreements from a financial reporting perspective, but the focus of this paper will be on the tax implications for the individual donor to a split-interest entity, especially as it relates to the new QCD provision.

The remainder of the paper will be organized as follows. A review of QCD provisions will be provided. That review will be followed by an explanation of the relevant split-interest entities that can receive QCDs under the new law. However, the discussion of these split-interest entities will be a general description of these entities as charitable gift vehicles rather than of the specific rules that apply to QCDs made to them. Then the new law allowing QCDs to these split-interest entities will be outlined, specifically discussing how QCDs can work with respect to these entities. Prior to the conclusion, several implications about the new QCD provision for transfers to split-interest entities will be noted. Interestingly, these implications, taken together, may mean that few charities will be emphasizing this new opportunity and that few individuals will use this new law for charitable giving.

QUALIFIED CHARITABLE DISTRIBUTIONS

A QCD is a direct transfer from an IRA to a qualifying charity. The amount transferred cannot be claimed by a taxpayer as an itemized charitable contribution deduction, but the amount is also never included in income. This provision was added by the Pension Protection Act of 2006 (P.L. 109-280). It was originally a temporary law but was extended multiple times. The Consolidated Appropriations Act of 2015 (P.L. 114-113) again extended this tax provision but did not include a sunset date, so the QCD currently has no expiration.

Since the QCD amount is never included in taxable income, a taxpayer can contribute to charity and get a tax benefit without itemizing. This possibility became even more important when the Tax Cuts and Jobs Act of 2017 (TCJA) (P.L. 115-97) substantially increased the standard deduction amounts for 2018-2025. The percentage of individual tax returns claiming the standard deduction was between 68% and 69% in 2016 and 2017, but after the TCJA was passed at the end of 2017, the percentage of returns claiming the standard deduction increased for 2018 through 2020 to more than 87% (IRS, SOI).

Qualifications for, limitations on, and benefits of QCDs have been described previously (Smith and Smith, 2017; Oestreich and Smith, 2022), but a review of the basic concepts will be mentioned, updated for the SECURE 2.0 Act of 2022 (P.L. 117-328, Division T, Sec. 307). The QCD must be a transfer directly from the IRA to the charity; it cannot first be withdrawn from the IRA and then sent by the IRA owner to the charity. It must come from an IRA, not another type of retirement account. However, it may be possible for other retirement accounts to first be rolled over into an IRA, after which a QCD could be made from the IRA. Most entities that qualify for charitable deduction status will qualify for a QCD, but donor-advised funds and private foundations cannot accept a QCD and maintain the tax benefits for the donor.

In cases where one is itemizing charitable deductions and the taxpayer receives something of value from the charity, the fair value of what is received must be excluded from the amount claimed as an itemized deduction. However, for a QCD, nothing of value can be received from the charity; the entire amount of the donation must otherwise be tax deductible.

The IRA account holder must be at least 70½ years old at the time of the transfer for it to qualify as a QCD. Even though the government allows many retirement accounts, including IRAs, to grow tax deferred for a period of time, the required minimum distribution (RMD) is the government's stipulation for taxpayers to make withdrawals from those accounts starting at a certain age that then become taxable and provide revenues to the government. RMD amounts are determined annually using life expectancies based on government mortality tables and the value of the IRA.

When the QCD was first legislated, this minimum age of 70½ was the same as the starting age for RMDs. However, the Setting Every Community Up for Retirement Enhancement (SECURE) Act of 2019, part of the Further Consolidated Appropriations Act, 2020 (P.L. 116-94), changed the RMD age to 72. More recently, the SECURE 2.0 Act of 2022 further increased the RMD age. For those who turn 72 after January 1, 2023, the age is increased to 73. For those born in 1960 or later, the RMD age is increased to 75 (P.L. 117-328). Even though the age for RMDs has increased, the minimum age for QCDs has not.

Since QCDs were first allowed, the annual maximum amount for a QCD has been $100,000 per taxpayer. [This limit can be reduced for a specific taxpayer who is over 70½, still earning compensation income, and making deductible contributions to an IRA account, as it might prevent a double tax benefit of deducting an IRA contribution and then using it for a nontaxable QCD.] If spouses both have IRAs and qualify, a joint return could potentially exclude up to $200,000, assuming each spouse transfers $100,000 through a QCD. The SECURE 2.0 Act of 2022 (P.L. 117-328, Division T, Sec. 307) now provides an indexing provision for this amount, so starting in 2024, the annual limit will increase over time.

Any amount transferred from an IRA, either as a direct withdrawal that becomes taxable, or as a QCD, counts toward the annual RMD, assuming the taxpayer is old enough to qualify for RMDs. [Technically, a QCD can come from a Roth IRA. However, because qualified distributions from Roth IRAs are already tax free and the original Roth IRA owner has no RMDs while alive, it would not make much sense to use a Roth IRA for a QCD.] However, any excess amounts withdrawn from an IRA or transferred as a QCD do not reduce future years' RMD percentages.

SPLIT-INTEREST ENTITIES

The new tax provision allowing a QCD to a split-interest entity only applies to certain entities: a charitable remainder annuity trust, a charitable remainder unitrust, or a charitable gift annuity. Prior to describing QCDs to split-interest entities in the following section, this section will introduce these split-interest entities in an overall context of charitable giving.

In general, these types of charitable gifts operate by having an individual create a trust, often through a trust company, or make a gift to transfer assets to a charity with the provision that the trust or charity will provide income from the assets to a beneficiary for a period of time, perhaps the remaining life of the donor (the lead beneficiary), with the charity claiming the remainder interest at the end of the life of the agreement. This may be an effective way to benefit a charity with assets not intended to be passed to heirs but when income from the assets is still needed or desired during the remaining life of the donor. These gifts can also qualify for a partial immediate charitable gift deduction for those itemizing deductions.

As the three qualifying split-interest entities are different from one another, they will each be explained in further detail. This section will also briefly explain the possible charitable tax deduction that can be available by using one of these options. Some concrete examples of when split-interest entities may be used for charitable gifts are also illustrated.

Charitable Remainder Annuity Trusts

A charitable remainder annuity trust is one where the donor transfers assets into an irrevocable trust from which fixed annuity payments are made at least annually to one or more beneficiaries. These beneficiaries must include at least one that is not a charity under Internal Revenue Code (IRC) section 170(c), and if any of the beneficiaries are individuals, they must be living when the trust is created. The life of the trust can be a fixed term in years (not to exceed 20) or be the life or lives of the individual or individuals named as beneficiaries, even if more than 20 years. The fixed annuity payment cannot be less than 5% of the initial net fair market value of the assets contributed to the trust, nor can it be more than 50% (IRC, section 664(d)(1)).

When the life of the trust ends based on the fixed term in years or the death of the beneficiary(ies), the remainder interest is transferred to a charitable organization as defined in IRC section 170(c). The value of the remainder interest must be at least 10 percent of the initial fair market value of assets contributed to the trust (IRC, section 664(d)(1)). This required percentage would be verified with actuarial calculations at the time the trust is created.

Trust distributions to the non-charitable beneficiary(ies) are taxed as ordinary income to the extent the trust has ordinary income for the current year or has prior-year undistributed ordinary income. Similarly, if trust ordinary income is all distributed, payments can be taxed, in order, as capital gain, other income, or a non-taxable return of corpus (IRS, Charitable Remainder Trusts).

Capital gains may arise if appreciated assets are contributed to the trust and later sold by the trust. Contributing appreciated assets may benefit the donor, as the fair value will be measured at the inception of the trust and be used in valuing the lead beneficiary payments, but the capital gain is not taxable to the donor at that time. Instead, the tax basis of the assets on the donor's books is transferred to the trust. Similarly, if the trust reinvests in capital assets that result in capital gains later, those may also lead to distributions that may be taxed to lead beneficiaries as capital gains.

The costs of administering the trust will be included in calculating trust income. These could include the costs of setting up the trust, managing the assets over time, making distributions, and filing tax returns and tax forms. Because of these costs, it may be unlikely that trust companies will want to set up these trusts with assets valued at less than $100,000, and, of course, they would prefer hundreds of thousands or millions of dollars of assets.

Split-interest trusts do not pay taxes (unless there is unrelated business taxable income) but do file information tax returns (Form 5227). These returns will report the financial activities of the trust, including accounting for the distributions from the trust, characterized as ordinary income, capital gain, other income, or a return of trust principal. The trust will provide each beneficiary with a Schedule K-1 (Form 1041) so these amounts can be correctly reported on the tax returns filed by the beneficiary(ies) (IRS, Charitable Remainder Trusts).

This type of trust would not allow additional contributions over time, as the income to the beneficiary(ies) is based on the fair value at the time of the initial donation. However, another trust could be created later for an additional contribution.

Charitable Remainder Unitrusts

A charitable remainder unitrust is like a charitable remainder annuity trust described above except that the payments to the lead beneficiaries are a fixed percentage (between 5 and 50 percent) of the net fair market value of the assets, valued annually. Thus, whereas a charitable remainder annuity trust has fixed dollar payments to lead beneficiaries, a charitable remainder unitrust has variable payments as the value of the trust increases (or decreases) over time (IRC, section 664(d)(2)).

A charitable remainder unitrust can be one of four different types:

1. A “standard unitrust,” which pays the stated percentage of fair value as valued annually regardless of trust income.
2. A “net income unitrust,” which pays the stated percentage of fair value as valued annually to the extent of trust income. This type will not pay out trust principal.
3. A “net income with makeup unitrust,” which pays the stated percentage of fair value as valued annually to the extent of trust income. If the payout is less than the stated percentage because it was limited by trust income, this amount can be made up in future years. This type will not pay out trust principal but can pay out additional income in future years if the payout had been limited by income in one or more prior years.
4. A “flip unitrust” is a net income unitrust that becomes a standard unitrust at a specified date or when a specific event occurs (Philanthropies).

Because the payout to the beneficiary(ies) is a percentage of the trust fair value, the trust can be set up to allow additional contributions over time. However, additional contributions to a charitable remainder unitrust must maintain the stipulation that the value of the remainder interest must be at least 10 percent of the fair value at the date of contribution (IRC section 664(d)(4)). If additional assets are added to the trust, the income to the beneficiary(ies) could increase as the value of the trust increases.

Charitable Gift Annuities

A charitable gift annuity involves the transfer of assets to a charity. In turn, the charity promises to make fixed annuity payments for life on at least an annual basis to one or two beneficiaries. While this arrangement does not require the use of a trust, the charity enters into a contractual agreement to guarantee these payments. The annuity amount is calculated such that the value of the contributed assets is larger than the actuarial value of the annuity, providing a remainder interest to be used for the charitable purposes of the charity.

IRC section 72 allows for some exclusion from gross income for annuity payments received as long as the total exclusion does not exceed the “unrecovered investment in the contract.” A charitable gift annuity is treated as if the donor has both (1) purchased an annuity, and (2) made a charitable gift. Thus, some portion of the annuity payments received from a charitable gift annuity may not be taxable. In addition, if appreciated assets are part of the amount contributed, the fair value will be used to determine the annuity payments made to the beneficiary(ies), but the capital gain can be recognized by the donor over the donor's expected life. Thus, some of the annuity payments may qualify for lower capital gain tax rates.

Depending on state law, the charity may be able to spend the entire gift (or at least the remainder interest) immediately. However, the charity still needs to make provisions to guarantee payments for the promised annuity.

For a charitable gift annuity, because the contributed assets have been transferred to the charity rather than to a trust, the costs of administering and reporting would be borne by the charity in its own financial statements. While there are still costs of administering a charitable gift annuity, they would be less than the costs of administering a trust. Therefore, charities may be more willing to accept assets worth $50,000, or even potentially less, for a charitable gift annuity. It may be possible that smaller charities will use other nonprofit entities or a charitable gift administration company to administer these gifts and make the annuity payments to beneficiaries, transferring the remainder, net of administration costs, to the charity. This way a small charity would not need to worry about managing this type of gift but could still accept them.

Because the annuity amount is calculated based on the fair market value of the original contribution, amounts would not be added to a charitable gift annuity once established. However, an additional contribution could be made as a separate charitable gift annuity.

Charitable Contribution Deduction

Because each of these split-interest entities leaves a remainder interest to a charity, a charitable contribution deduction is allowed for a portion of the amount of the donation. The deduction allowed is based on IRS formulas, which include the age of the donor and the income beneficiaries, the payout of the trust or annuity, and an IRS rate called the applicable federal rate (Philanthropies).

However, limits exist for a charitable contribution deduction based on a taxpayer's adjusted gross income (AGI), so these limits need to be considered. In many cases, the limit is 60 percent of AGI, but there are also 20 percent, 30 percent, and 50 percent limits in certain cases. Amounts that exceed the limits can be carried forward for up to five future tax years and potentially provide a tax benefit in those years.

The TCJA (P.L. 115-97) made several changes to itemized deductions, eliminating some previous deductions and further limiting others, but it also eliminated the "Pease limitation," which reduced the ability to claim itemized deductions for higher-income taxpayers. This limitation, with indexing and other changes, was in place from 1991 to 2009 and again from 2013 to 2017. While the TCJA eliminated this limitation on itemized deductions for 2018-2025, it could return, depending on what happens to the TCJA provisions by the time they will otherwise expire.

Examples

The following three examples illustrate situations where income for life with a charitable remainder might be desirable. A split-interest entity may be helpful in achieving these goals.

Example 1: Ellen is a wealthy widow who has been involved with and supportive of a specific charity for many years and wants to leave a legacy to the charity when she passes. She has already provided for her heirs as she desires. She may choose to put her remaining assets into one of these split-interest entities that will provide adequate income for the remainder of her life, leaving the balance to the charity upon her passing.

Example 2: Carlos and Maria are aging parents of one son, Mateo, who is 60 years old. Mateo is disabled, has never married and has no children, and is unable to provide for himself. Carlos and Maria anticipate dying before Mateo but want to provide support for Mateo for the remainder of his life. As there are no other heirs, they want the assets left after Mateo dies to go to a charitable entity researching ways to prevent and treat Mateo's disability. Carlos and Maria may choose to transfer assets to a split-interest entity that will provide support to Mateo for the remainder of his life and provide resources to the research charity upon Mateo's passing.

Example 3: John and Evelyn have two children, Kaylyn and Jody, both of whom they want to consider in their estate plans. John and Evelyn consider that Kaylyn is financially responsible, and they intend to leave some assets to her, as they feel she can appropriately manage the assets and use them wisely according to her needs. However, Jody has proven to be irresponsible with her resources and does not seem to have the internal motivation to save and spend with care. So rather than leave assets to Jody that she may squander quickly, John and Evelyn may choose to put assets into one of these split-interest entities to provide Jody income systematically over the remainder of her life, thus reducing the risk that she will be able to waste substantial amounts in a short time.

There may also be many other similar situations that would justify using a split-interest entity. The SECURE 2.0 Act (P.L. 117-328, Division T, Sec. 307) now allows a QCD as a vehicle to fund any of the three types of split-interest entities discussed above. However, this new law has specific provisions or limitations that preclude application to Examples 2 and 3 and limit application to Example 1. A description of this law is provided, followed by a discussion of the implications.

QCDs TO SPLIT-INTEREST ENTITIES

IRC section 408(d)(8) details the requirements for a QCD. The SECURE 2.0 Act of 2022 (P.L. 117-328, Division T, Sec. 307) amended IRC section 408(d)(8) by adding a new possibility for a QCD. On a one-time basis, a taxpayer can now make a transfer from an IRA to a split-interest entity as a QCD. For the purpose of this QCD, a split-interest entity can include a charitable remainder annuity trust, a charitable remainder unitrust, or a charitable gift annuity. These entities are defined in IRC sections 664(d)(1), 664(d)(2), and 501(m)(5), respectively, and were described earlier. For any of these entities to qualify for a QCD, each must be funded exclusively by QCDs. In addition, in the case of a charitable gift annuity, fixed annuity payments of 5 percent or greater must commence within one year of the date of the funding. This 5 percent requirement for charitable gift annuities, in the case of a QCD, makes them similar to charitable remainder annuity trusts and charitable remainder unitrusts, which already require at least a 5 percent payout rate.

Although QCDs cannot qualify as charitable deductions, for any of these split-interest options for a QCD the value of the remainder interest for the charity would otherwise need to be allowable under IRC section 170 as a charitable deduction. In addition, no person can hold "an income interest in the split-interest entity other than the individual for whose benefit such account is maintained, the spouse of such individual, or both." The income interest in the split-interest entity also cannot be assignable (P.L. 117-328, Division T, Sec. 307).

Other special rules also apply to this type of QCD. For either of the trust options, the payments to the non-charitable beneficiary(ies) will be treated as ordinary income. Thus, these payments cannot have capital gain status or tax-free status to the recipients. QCDs used for a charitable gift annuity are not treated as investments in the contract for purposes of IRC section 72. Therefore, none of the annuity payments received from a charitable gift funded by a QCD can be excludable and will all be treated as ordinary income (P.L. 117-328, Division T, Sec. 307).

The maximum amount for this new QCD is $50,000, but the amount will be indexed starting in 2024. Apparently, the "one-time" election means that the election can only be made in one tax year, although it could be done through several QCD transfers in that tax year, at least to a charitable remainder unitrust that can allow additional contributions, as long as the aggregate amount does not exceed the maximum limit (P.L. 117-328, Division T, Sec. 307).

IMPLICATIONS

Perhaps the most important implication from this new law allowing a one-time QCD to fund a split-interest entity is the tax advantage to the donor. A comparison of funding a split-interest entity with cash from an IRA versus funding it through a QCD illustrates the advantage.

If cash is withdrawn from an IRA for purposes of contributing to a split-interest entity, the amount withdrawn becomes taxable to the account owner. The donor can then receive a charitable contribution deduction for a portion of the contribution to a split-interest entity. However, the benefit can only be obtained if the donor is itemizing and is not subject to charitable contribution deduction limits, either for a percentage of AGI or if the "Pease limitation" is reinstated in the future and the donor has high enough income to lose some of the benefit of itemizing.

On the other hand, if the new QCD is used to contribute to a split-interest entity, although there is no tax deduction available for the calculated remainder interest to the charity, there is no tax liability for the amount withdrawn from the IRA. In this case, the donor gets a tax advantage (not paying tax on the IRA distributions) for the entire amount rather than just for the partial amount (the amount allowed as an itemized deduction), and the taxpayer does not need to itemize to get this advantage. In addition, the amount withdrawn from the IRA can still qualify as part or all of the annual RMD for that year. Also, the amount transferred never becomes part of AGI. This can result in less Social Security income being taxable, lower Medicare Part B and Part D premiums two years following, or other possible benefits of lower AGI (Oestreich and Smith, 2022).

However, since the trust or charitable gift annuity payments from the QCD contribution to the non-charitable beneficiary will be defined as ordinary income, there will also be no potential tax advantage for the life income. For example, if appreciated assets are contributed to a charitable remainder trust, some of the income to the lead beneficiary could receive capital gain treatment. If appreciated assets are contributed to a charitable gift annuity, the capital gain can be recognized over time by the donor/beneficiary. Likewise, the calculated value of the amount invested in the annuity can be received by the annuitant as nontaxable income. These advantages are not available if the split-interest entity is created through a QCD.

The requirement for lead beneficiary income from one of these split-interest entities created through a QCD to be ordinary income may stem from an assumption that the QCD transfer from the IRA to the trust or charity will be a transfer of cash, not appreciated assets. If any of the assets sold in the IRA to get the cash to transfer were appreciated assets, those gains are not taxed while in the IRA anyway, or if they are considered to be part of the cash transferred in the QCD, they are not taxed because of the exclusion from income for the QCD transfer. Therefore, if an additional tax benefit of qualifying some of the payments from the trust or annuity as capital gain or a return of investment were allowed, that could be considered a double tax benefit.

In addition, while ordinary income may not have the same tax advantages, defining the beneficiary income as ordinary income might make income calculations and reporting easier for the trust or charity. It might also make tax reporting for the beneficiary easier, as the income would not be split into different types of taxable (or nontaxable) income.
A second implication is that the split-interest entity can only come from QCDs. In many cases, this would mean the maximum amount would be $50,000 (indexed). Even if Ellen in the first example above had all her remaining assets in an IRA, she could only make a QCD from the IRA to a split-interest entity for $50,000. She would need to consider another split-interest entity for the remainder of her estate to provide life income with the remainder left to the charity. It is possible that trust companies may not want to set up and administer a trust for such a relatively small amount as $50,000, although she might be able to use a charitable gift annuity.

If both spouses have IRAs and transfer $50,000 each through a QCD, it is possible the total contribution to the split-interest entity could be $100,000 (indexed). Many charities would accept this amount for a charitable gift annuity. But even with a $100,000 original contribution, trust companies might be hesitant to set up and administer a separate trust for this amount. The costs for setting up and administering a trust may necessitate larger asset transfers to make a trust worthwhile. Also, if two spouses were each making $50,000 QCD transfers to the same split-interest entity, they would need to make the transfers at the same time for a charitable remainder annuity trust or a charitable gift annuity, as these entities can only accept contributions at inception.

A third implication focuses on the government loss in revenues by giving this enhanced tax break for a QCD to a split-interest entity. Of course, every time the government enacts a tax break for taxpayers, it simultaneously gives up possible revenues to fund government services. Since many tax laws are enacted to promote social or economic actions that may benefit society, the tradeoff between any increased benefits to charities and the reduced revenue to the government needs to be considered. Perhaps this unknown tradeoff may be the reason for the current law allowing only a one-time QCD to a split-interest entity and for limiting the dollar amount. Is it possible that either the restriction to a one-time QCD or the $50,000 limit might be relaxed in the future?

A fourth implication relates to the one-time QCD to a split-interest entity. Although it might be possible to make multiple QCD transfers, they would all have to be made in one year and could not exceed the $50,000 (indexed) limit. While multiple QCD transfers in one year might work for a charitable remainder trust which can accept additional contributions, they would not work for a charitable remainder annuity trust or a charitable gift annuity, as they cannot accept additional amounts. Because charities would not likely encourage QCDs to split-interest entities for less than $50,000, the possibility of making multiple QCD transfers in one year likely has little benefit.

A fifth implication relates to the requirement that the income beneficiary(ies) of a split-interest entity from a QCD must be only the IRA account holder, a spouse, or both. This means that Carlos and Maria from the second trust example and John and Evelyn from the third trust example could not use a QCD for any portion of the benefits they want to provide for Mateo and Jody, respectively. Since there are already special rules about RMDs for non-spouse beneficiaries of the IRA, the prohibition of using a QCD for a beneficiary other than the account owner or spouse may be intended to keep people from avoiding these IRA beneficiary rules by using a split-interest entity that may provide income for life to the beneficiary.

This implication also means there cannot be more than two lead beneficiaries of a trust funded through a QCD. While a charitable remainder annuity trust or a charitable remainder unitrust can, in general, have more than two lead beneficiaries, this flexibility does not apply to trusts created by QCDs. Charitable gift annuities are already limited to not more than two lead beneficiaries; for a charitable gift annuity created through a QCD, the number of lead beneficiaries is not further limited, but they can only be the IRA owner, a spouse, or both.

A sixth implication that already applies to QCDs is that one must have an IRA (or other accounts that can be converted to an IRA) to make a QCD. If an individual’s retirement income comes only from a defined benefit pension plan or from other nonqualified investments, no QCD would be possible for that individual.

A seventh implication relates to the requirement for the split-interest entity to pay out at least 5 percent to the non-charitable beneficiary(ies). The American Council on Gift Annuities (ACGA) develops and publishes suggested maximum gift annuity rates. The ACGA claims that that its standards are followed by 97 percent of charitable organizations nationwide (acga-web.org). The suggested rates are updated periodically with the latest rates effective as of January 1, 2023. If a QCD is transferred to a charitable gift annuity with the IRA account holder as the single-life annuitant, the 5 percent minimum should not be a problem, as the donor cannot use a QCD until at least age 70½, and the current suggested maximum gift annuity rate for someone age 70 is 5.9 percent (with a 5.0 percent maximum rate starting at age 61) (acga-web.org, current rates).

On the other hand, if the IRA account holder is old enough to make a QCD but the spouse, as the single-life annuitant of a charitable gift annuity is significantly younger, that might not qualify. The spouse would need to be at least 61 years old to make the current suggested maximum gift annuity rate 5 percent or higher. If the spouse is younger than 61, the QCD to such a charitable gift annuity may need to be postponed.

If a QCD is used to fund a charitable gift annuity with both spouses as beneficiaries of a two-life survivor annuity, the maximum suggested rate may not be as high as 5 percent if the younger spouse is significantly younger than the spouse using the QCD. Even though the individual making the QCD would have to be at least 70½, a spouse much younger than the IRA holder making the QCD may not qualify the couple for a 5 percent annuity rate. For example, if the younger spouse is 62 or younger, there is no age for the older spouse that will qualify the couple for a 5 percent rate (or higher) under the current suggested maximum gift annuity rates. In this case, the QCD may need to be postponed until both spouses are old enough for their joint annuity to reach the 5 percent minimum requirement. Table 1 shows other ages for which a couple could qualify for at least a 5 percent annuity rate using the current suggested maximum gift annuity rates where the older spouse would be old enough to make a QCD (acga-web.org, current rates).

Table 1
Ages for Spouses to Qualify for Suggested Minimum Gift Annuity Rates of at Least 5 Percent
Two Lives—Joint and Survivor
(ACGA Rates Effective January 1, 2023)

Age of Younger Spouse	Minimum Age for Older Spouse
63	76
64	73
65	71
66	70

An eighth implication is that a deferred charitable gift annuity cannot qualify for a QCD if the deferral period is more than one year. Any annuity from this QCD would need to start immediately or within one year.

A ninth implication is that, although a QCD normally does not allow the donor to receive any benefits from the charity, a QCD to a split-interest entity certainly involves the donor (and/or spouse) receiving benefits from the trust or charitable gift annuity. However, it is still true that the value of the remainder interest must otherwise be allowable as a charitable deduction under IRC section 170.

A tenth implication arises because a charitable remainder annuity trust or a charitable remainder unitrust may have the donor as the trustee. This is true for these trusts in general and can be true for a trust created by a QCD transfer. However, as the requirements for setting up and administering such a trust can be quite complicated, this may not be the best choice in many cases. If a donor is sophisticated in these areas, it might make sense for a donor to become the trustee to save administrative costs that might otherwise be charged by a trust company, but it might also be risky. In addition, if a charitable remainder trust names the donor as a life-income beneficiary, what would happen when the donor dies? A successor trustee would need to be named to transfer the remaining trust assets to the charity. Otherwise, this might result in a costly legal hassle.

An eleventh implication is also true of charitable gifts in general and would apply to charitable remainder trusts and charitable gift annuities, regardless of whether they were created through a QCD transfer or created otherwise. The donor can include permanent restrictions that the gift is to be invested and the income used by the charity. An example might be an endowment for a cancer research program. The principal donated cannot be spent but will be invested, and the earnings from the investments can be used by the charity for cancer research. Temporary restrictions can also be stipulated by the donor; these would be purpose or time restrictions. For example, a donor may want to support her alma mater but wants to specifically support the department from which she graduated. So, rather than simply making a charitable gift to the university, it may be made to the university with a purpose restriction that it be used for scholarships in the dance department. A time restriction may be a stipulation by a donor that the gift cannot be used by the charity for a specific period of time. For example, a charitable remainder annuity trust may specify that the annuity payments to the lead beneficiary(ies) are to be made for ten years and that the charity's remainder interest cannot be spent by the charity until five years after that.

The potential increased tax advantage of a QCD to a split-interest entity compared to contributing taxed cash and taking an itemized deduction makes this new tax provision sound quite helpful, both to charities and to those with IRAs who desire to support charities. However, taken together, these implications may make it difficult for charities and trust companies to focus a lot of effort in pursuing seniors who may want to make one-time QCD transfers to a split-interest entity. The administrative costs for the trusts, both initially and over time, may be too high to make this provision of much help to charities. Since the administrative costs for a charitable gift annuity are less than for a trust, a one-time QCD of $50,000 to a charitable gift annuity may be more likely to be acceptable to a charity.

The implications may also make it difficult for seniors to give serious consideration to this tax provision. It may not provide much life income based on the dollar limitation of the QCD and the limit for a one-time-only transfer. For example, a $50,000 charitable gift annuity with a 5 percent annuity payment would only result in a $2,500 payment to the beneficiary(ies) annually. If this small amount of annual income is not really needed, it might be better to just make a QCD directly to the charity rather than to a split-interest entity. That way the charity would get the full benefit of the QCD amount rather than just the remainder interest. A QCD to a split-interest entity may also not be very helpful to those who want to make exceptionally large charitable gifts with a significant amount of lifetime income, as it would only be one small piece of a plan to support a charity with these large gifts.

CONCLUSION

While the main purpose of the federal income tax is to raise appropriate resources to help fund government operations, tax policy is often used to encourage specific social or economic behaviors. A major focus of the SECURE 2.0 Act was to make changes to encourage retirement savings, both by motivating employers to offer such plans and by providing opportunities for employees to benefit by participating in such plans. However, it also included a provision for a new one-time QCD to a split-interest entity, which also targets charitable giving. Time will tell if this new provision will have the intended effect of increasing donations to split-interest entities.

The dollar limit for this QCD, even though indexed for future years, may not be adequate to motivate many qualified seniors to take advantage of this opportunity. It may also be inadequate to motivate trust companies and charities to work with amounts this small, especially for very wealthy donors who might want to set up charitable remainder trusts or charitable gift annuities in the amount of multi-millions of dollars. If the concept of this QCD is well received but not acted upon by many donors, the dollar limit may need to be significantly increased in the future if the government really wants to encourage this type of charitable giving by seniors and if this tax policy is the best way to motivate this behavior.

REFERENCES

Accounting Standards Codification (ASC). Financial Accounting Standards Board. fasb.org

American Council on Gift Annuities, https://www.acga-web.org.

American Council on Gift Annuities, https://www.acga-web.org/current-gift-annuity-rates.

Internal Revenue Code of 1986 (IRC), Title 26, U.S. Code, as amended.

Internal Revenue Service, Charitable Remainder Trusts, https://www.irs.gov/charities-non-profits/charitable-remainder-trusts.

Internal Revenue Service, Statistics of Income (SOI), Individual Income Tax Returns Complete Report, 2020, Table A, Publication 1304, Rev. 11-2022, https://www.irs.gov/pub/irs-pdf/p1304.pdf.

Oestreich, Nathan and Sheldon R. Smith (2022). "Qualified Charitable Distributions Four Plus Years After Tax Reform," Decision Sciences Institute, 53rd Annual Conference Proceedings, November, 862-868.

https://philanthropies.churchofjesuschrist.org/gift-planning/how-to-give/split-interest-giving-tools/

Public Law 109-280. Pension Protection Act of 2006, August 17, 2006.

Public Law 114-113. Consolidated Appropriations Act, 2016, December 18, 2015.

Public Law 115-97. Tax Cuts and Jobs Act of 2017, December 22, 2017.

Public Law 116-94. Further Consolidated Appropriations Act, 2020, including Setting Every Community Up for Retirement Enhancement [SECURE] Act of 2019, December 20, 2019.

Public Law 117-328. Consolidated Appropriations Act of 2023, including SECURE 2.0 Act of 2022, December 29, 2022.

Smith, Sheldon R. and Lynn R. Smith (2017). "Qualified Charitable Distributions from an IRA," Western Decision Sciences Institute, Annual Conference Proceedings, April.

DECISION SCIENCES INSTITUTE

Risk Quantification of US Sectors Around the Pandemic

Mehmet Orhan
University of North Texas
Email: Mehmet.Orhan@unt.edu

Levent Bulut
University of North Texas
Email: Levent.Bulut@unt.edu

ABSTRACT

We explore the risk aspects of US sectors around the pandemic. The volatility quantified by the variance of stock market index series of the sectors are summarized to observe the behavior of risk before and during the pandemic. We make use of the most widely employed measure of risk, VaR, to check if it was able to quantify risk sufficiently by backtesting with methods introduced by Kupiec and Christoffersen. Results reveal that the hit of the pandemic had a substantial impact on the quantification of risk by VaR. VaR experienced the highest difficulty in measuring the risk of Consumer Staples sector.

KEYWORDS: Value-at-Risk, Sectors, GARCH(1,1), Backtesting, Kupiec Test, Christoffersen Test

INTRODUCTION

Banks, corporations, and regulators allocate resources to estimate the exposure and potential loss of a portfolio to unfavorable market movements. There are several motivations behind these calculations of potential loss, including but not limited to, managing and mitigating risk, facilitating better decision making, ensuring regulatory compliance, performing stress testing, optimizing portfolio allocation, and evaluating performance.

Globalization of financial markets facilitated spread of crises at any country to the entire world rapidly. When COVID-19 was first observed in Wuhan, China, on December 31, 2019, no one could estimate its severe impact on financial markets. It started shaking the global markets after the World Health Organization (WHO) declared it a "Public Health Emergency of International Concern" on January 30, 2020, due to the extraordinary swift infection rate of the virus. Unlike many previous crises, COVID-19 did not originate from the financial sector, but its ripple effects have significantly impacted the financial world. According to the World Bank's 2022 World Development Report, COVID-19 pandemic triggered the largest global economic crisis in over a century. In 2020, economic activity contracted in 90 percent of countries, with world GDP declining by about 3 percent. The impact of this crisis on financial markets is unparalleled after the Great Depression. While it has been highly detrimental to financial markets, it also presents a unique opportunity to examine the risk dynamics of these markets. This paper is primarily motivated by the aim to make use of this opportunity to explore risk quantification using the most popular method, Value-at-Risk over the US market series.

Value-at-risk (VaR) is in fact a statistical technique which measures the maximum expected amount a portfolio might lose over a specified time period, most commonly one day.

VaR provides a quantitative measure of the potential loss in the value of a financial asset's portfolio. Some theoretical studies have pointed out the limitations of VaR due to its lack of sub-additivity and convexity (see (Cheng, Liu, and Wang, 2004), for example), but no better measure has been found to quantify risk. The method is based on the assumption that the future will be like the past. However, this presumption is not always valid. History has shown that volatility can vary over time, and that past returns are not always a good predictor of future returns. For instance, the financial crisis of 2008 was a major event that had a significant impact on the global financial markets. As usual, the crisis was not predicted by historical models, and it caused significant losses to many investors. Another limitation is that the historical method does not take into account changes in market conditions. Let's give a timely example. The current increase in interest rates due to aggressive contractionary monetary policies by the Fed lowers the value of assets that are sensitive to interest rates. The historical simulation approach which directly uses the historical series to forecast the VaR cannot predict these changes, and it may therefore underestimate the true level of risk. Likewise, the COVID pandemic was an unprecedented event that caused a global economic crisis and a sharp decline in the market returns. Any approach based on history would not be able to account for such a scenario and would underestimate the risk of future losses.

As a remedy, the literature emerged in using GARCH specification to estimate the VaR. Volatility is a measure of how much the price of an asset fluctuates over time and many time series have shown persistent clustering of volatility in the financial markets. One can argue that the volatility inherently contains information about the uncertainty or risk in a market. When volatility increases, it is expected that the price of the asset is more likely to change significantly in a short period of time. This can be due to a number of factors, including unforeseen events such as financial crises, interest rate changes, and pandemics. The logic is very simple: an unforeseen event triggers a sudden and significant increase in volatility. In response, investors become more uncertain about the future, and they are more likely to sell assets in order to reduce their risk. This selling pressure leads to sharp declines in prices. By identifying shifts in market dynamics through changes in volatility, we can potentially enhance the accuracy of VaR predictions, making the risk management process more reliable.

LITERATURE REVIEW

The ARCH/GARCH models, first brought to the academic scene by Engle in 1982 and subsequently by Bollerslev in 1986, have been developed in more recent years to address the issues related to clustering in financial data. (Bollerslev, 1986) extended Engle's work with the introduction of GARCH models, which integrated lagged conditional variances into the model, providing a more comprehensive method for capturing volatility dynamics.

These models condition on heteroskedasticity to monitor and control fluctuating volatility. (Christoffersen, 1998) and (Jorion, 2007) showed that ARCH/GARCH models are better at estimating VaR than other models because they can adapt to changing market conditions and capture the fat tails that are often seen in financial return distributions. This adaptability is important, especially during periods of financial distress. More recently, studies by (McAleer and Da Veiga, 2008) and (Hansen, 2011) have shown the effectiveness of ARCH/GARCH models in forecasting VaR in different markets and for various asset classes, including equity indices. In estimating Value at Risk (VaR), the work by (Orhan and Koksal, 2012) provides a useful comparison of various GARCH models.

Covid-19 pandemic offered an appropriate opportunity to benchmark the performance of methods to calculate VaR but research benefiting from this opportunity is very limited. (Ahadiat and Kesimah, 2021) used data from January 2020 to the fourth quarter of 2020 to analyze the potential loss from investing in the stock market of four government banks by revisiting value at risk (VaR) as a tool for measuring the maximum loss. They conclude that AR (1)-GARCH (1) is a good fit for the determination of the mean and variance model and recommend that investors who have funds in state-owned banks should reconsider their investments. Similarly, (Omari et al, 2020) combined conditional heteroscedastic models and extreme value theory to examine the tail behavior of stock indices from major economies over January, 1, 2006 to July, 31, 2020 and conclude that the ability of conditional extreme value theory to capture volatility clustering accurately is not sufficient for a correct assessment of risk in these markets. The research conducted by (Vlaar, 2000) presents an in-depth analysis of VaR models as applied to Dutch bond portfolios, shedding light on the performance of these models in a specific asset class and geographical context, thus enhancing our understanding of VaR estimation in different market settings. Last but not the least, (Shaik and Padmakumari, 2022) investigate the VaR estimation models and their predictive performance by applying a series of backtesting methods on BRICS (Brazil, Russia, India, China, South Africa) and US stock market indices. They use normal, historical simulation and exponential weighted moving average models over global financial crisis, and COVID-19 periods. Their results show that the best performer is the EMWA model and VaR models perform poorly during global financial crisis and COVID-19. They also conclude that the predictive accuracy of VaR methods is weak during the COVID-19 era.

This study offers important insights into the performance of the GARCH(1,1) model in VaR estimation, thereby contributing significantly to the literature on financial risk management. As the literature review suggests there is no study making use of data till the end of the pandemic and there is no research output on the comparison of risk behavior of US sectors. Our paper fills these two major gaps.

DATA SET

The data we're working with covers the period from May 31, 2012, to June 10, 2022. We use US sector indexes to proxy the impact of the pandemic on the sectors listed below:

- Communication Services, (CSer)
- Consumer Discretionary/Consumer Cyclical (CDis)
- Consumer Staples/Consumer Defensive (Csta)
- Energy, (Ener)
- Financials, (Fina)
- Health Care, (Heal)
- Industrials, (Indu)
- Information Technology, (Info)
- Materials, (Mate)
- Real Estate, (Real)
- Utilities, (Util)

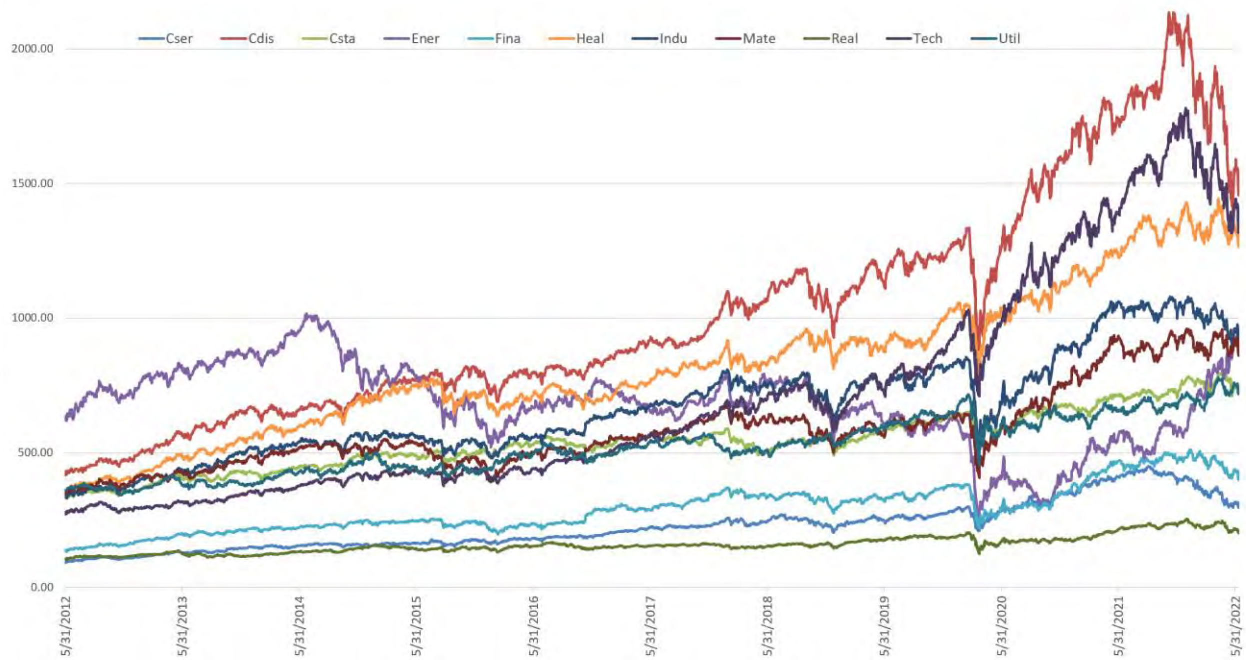
Plot of the closing prices for each day over the entire period is displayed in Figure 1. Note that the sector closing price series reveals the downturn of all markets when hit by the pandemic clearly. Such sudden movements are too difficult for even GARCH models to handle the volatility. Similarly, measures against the adverse impact of the pandemic embed surprises in the series.

Descriptive Statistics

We present the summary statistics of the returns to observe the snapshot of sector performances in Table 1. Median figures are substantially different than the means underlining the existence of the outliers. Portfolio investor who manages risk by refraining from highly volatile markets could have returned 0.126% on the average by simply investing in Consumer Staples/Consumer Defensive sector with the lowest variance of 0.816. The lowest volatility (proxied by the standard deviation), on the other hand belongs to Consumer Staples/Consumer Defensive. Energy sector surprises the typical investor with highest and lowest returns (max return of 16.30% and minimum return of -20.46%) leading to the highest range of 36.77%.

We allocated the last line to the coefficient of variation to highlight the volatility per unit return earned. By far, the highest CV is from the Energy sector followed by the Real Estate. To the contrary, lowest volatility per percentage return is from Health Care.

Figure 1: Closing Prices of all Sectors



All but Utilities markets have negative skewness indicating the distribution skewed to the left. All descriptive statistics are computed with Excel and Excel uses the excess kurtosis figure

Table 1: Summary Statistics of Sector Returns

	Cser	CDis	Csta	Ener	Fina	Heal	Indu	Mate	Real	Tech	Util
Mean	0.052	0.056	0.034	0.031	0.051	0.055	0.046	0.043	0.032	0.070	0.034
Median	0.080	0.126	0.048	0.035	0.078	0.087	0.078	0.079	0.093	0.095	0.092
Mode	0.000	#N/A	0.000	0.000	0.000	0.000	#N/A	0.000	0.000	#N/A	0.000
St Dev	1.180	1.199	0.903	1.800	1.370	1.053	1.203	1.260	1.204	1.336	1.142
Variance	1.392	1.437	0.816	3.241	1.876	1.108	1.448	1.588	1.449	1.786	1.303
Kurtosis	8.66	12.09	16.43	16.54	16.70	9.44	15.26	10.45	21.91	12.60	21.55
Skewness	-0.592	-0.803	-0.278	-0.444	-0.274	-0.253	-0.372	-0.385	-1.084	-0.308	0.082
Range	19.86	22.20	17.76	36.77	27.22	17.58	24.20	23.13	25.18	25.88	24.64

Minimum	-11.12	-12.74	-9.32	-20.46	-13.99	-9.99	-11.45	-11.51	-16.55	-13.92	-11.53
Maximum	8.75	9.46	8.44	16.30	13.23	7.59	12.75	11.61	8.63	11.96	13.12
Coef of Var	22.68	21.59	26.26	57.87	26.81	19.08	26.31	29.27	37.26	18.97	33.71

with a minimum of 8.66 revealing that the return series have thick tails for all markets. Kurtosis coefficients are all significantly higher than 3 telling us that the series are leptokurtic with thick tails. This is not surprising as the series are having high volatility.

The correlation matrix in Table 2 provides an insight into the relationship between the returns of different sector indexes. There is a strong positive correlation between most of the sector pairs. For instance, the Commercial Services sector shows high positive correlation with Consumer Discretionary, Consumer Staples, Financials, Health Care, Industrials, Materials, Real Estate, Information Technology, and Utilities sectors, with correlation coefficients exceeding 0.9. Conversely, the Energy sector displays a strong negative correlation with all other sectors, with correlation coefficients ranging from -0.344 to -0.573. There is a nearly perfect positive correlation (0.988) between the Financials and Industrials. These correlations suggest that most sectors tend to move in the same direction, with the notable exception of the Energy sector, which often moves in the opposite direction.

Table 2: Correlation between Sector Index Pairs

	Cser	CDis	Csta	Ener	Fina	Heal	Indu	Mate	Real	Tech	Util
Cser	1.000	0.987	0.939	-0.573	0.945	0.967	0.971	0.950	0.924	0.972	0.917
CDis	0.987	1.000	0.949	-0.551	0.943	0.978	0.967	0.957	0.941	0.987	0.925
Csta	0.939	0.949	1.000	-0.532	0.925	0.975	0.950	0.925	0.958	0.927	0.965
Ener	-0.573	-0.551	-0.532	1.000	-0.367	-0.527	-0.433	-0.344	-0.452	-0.549	-0.553
Fina	0.945	0.943	0.925	-0.367	1.000	0.946	0.988	0.958	0.930	0.913	0.908
Heal	0.967	0.978	0.975	-0.527	0.946	1.000	0.967	0.949	0.952	0.960	0.950
Indu	0.971	0.967	0.950	-0.433	0.988	0.967	1.000	0.971	0.937	0.937	0.933
Mate	0.950	0.957	0.925	-0.344	0.958	0.949	0.971	1.000	0.922	0.951	0.883
Real	0.924	0.941	0.958	-0.452	0.930	0.952	0.937	0.922	1.000	0.927	0.948
Tech	0.972	0.987	0.927	-0.549	0.913	0.960	0.937	0.951	0.927	1.000	0.906
Util	0.917	0.925	0.965	-0.553	0.908	0.950	0.933	0.883	0.948	0.906	1.000

VaR MODEL

We have selected the most successful method of VaR calculation to the best of our knowledge which uses the GARCH(1,1) process. This model is a popular choice for VaR calculation because it can account for volatility clustering, tendency for volatility to be higher during periods of high volatility. To calculate VaR using GARCH (1,1), we make use of parametric method to compute VaR with each market sector series. Then, we use the model to predict how much the values in the series are likely to vary. We then get the VaR by taking this predicted variation and multiplying it by a factor that represents the likelihood of loss. More technically, the VaR calculation with GARCH(1,1) can be summarized as follows:

$$R_t = \mu + \epsilon_t \tag{1}$$

$$\epsilon_t = \sigma_t \, \upsilon_t \tag{2}$$

$$\sigma_t^2 = \omega + \eta \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3}$$

where $R_t = 100\,\frac{P - P_{t-1}}{P_{t-1}}.P_t$ represents the closing price of the sector index at the close of day t, μ is the constant mean of returns, υ_t is a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and variance 1. We assume that υ_t follows the standard normal distribution. Value-at-Risk, VaR(1- α), is defined as:

$$Pr(R_t < VaR(1-\alpha)) = \alpha \tag{4}$$

The computation of VaR depends on the variance term as:

$$VaR(1-\alpha) = \bar{R} - \hat{\sigma}_t z_\alpha \tag{5}$$

where $\bar{R}$ is the mean return and z_α is the critical value of the standard normal distribution with right-tail area α.

Literally VaR(1- α) is a value to be exceeded 100*α times every 100 days. It is a very useful, functional and simple definition, but the computation of VaR is challenging. Among the several methods of VaR computation like historical simulation, parametric VaR with GARCH(1,1) using standard normal critical values is the best performer ever. Computation of the variance with GARCH specification is just right for our series since all sector return series are proven to have the ARCH effect.

We use the 1000 days data to predict the standard deviation of the 1001st day on a rolling basis using equations (1) - (3), and in turn calculate the VaR with (4) and (5). Despite the straightforward conceptual basis of VaR for risk measurement, its computation involves complex procedures. We adopt the GARCH(1,1) method as suggested by (Orhan and Koksal, 2012) to compute VaR at confidence levels of 90%, 95%, and 99%.

We utilize data that predates the onset of the pandemic, extending our analysis up until the point when the prevalence of COVID-19 began to decline, signifying the initial phase of normalization. Since we also aim to explore the impact of the pandemic on US sectors, we make use of data from US market indices. More specifically, we are concerned with VaR's ability to quantify risk, even when financial markets are abruptly hit by an exogenous and unexpected shock.

BACKTESTING

To ascertain risk, we utilize backtesting as the primary criterion. For this purpose, both (Kupiec, 1995) and (Christoffersen, et al 2001) tests are employed. Kupiec's test is used to verify if the number of exceptions (i.e., instances when the actual loss exceeds the VaR) is consistent with the confidence level used in the VaR model. Kupiec's null hypothesis for number of violations is H_0: V/N = α where V is the number of violations, and thus the proportion of failure test statistics is:

$$K = 2\,ln\left(\frac{\left(1-\frac{V}{N}\right)^{N-V}\left(\frac{V}{N}\right)^{V}}{(1-\alpha)^{N-V}\,\alpha^{V}}\right) \tag{6}$$

Ideally percentage of violations for VaR(1- α) should be 100*α. That is, if the constructed model worked perfectly then "V in %" column in Table 3 for VaR(99%) should be 1 for all sectors (similarly it should be 5 for VaR(0.95), and 10 for VaR(0.90)). Test statistic assumes higher values as this percentage deviates from 1 in both directions. The null hypothesis asserting the perfect performance of the model is rejected for all series but Energy for VaR(99%). One can

reject H_0 for also Energy sector with exact probability of Type 1 Error equal to 3.7% which is pretty reasonable.
Indeed, GARCH model performance is slightly better for VaR(95%) with two p-values higher than 5%. And the best performance is by VaR(90%). The null is failed to be rejected for Communication Services, Consumer Discretionary/Consumer Cyclical, Financials, Industrials, Materials and Information Technology safely.

One can predict the favorable performance of GARCH by simply looking at the violation percentages around 10 for VaR(0.90). This percentage is within 10% interval for 6 of the sectors and is close to this borderline for 3 sector return series.

Table 3: Kupiec Test Results

	VaR(0.99)				VaR(0.95)				VaR(0.90)			
	V	V in %	Test Stat	p-val.	V	V in %	Test Stat	p-val.	V	V in %	Test Stat	p-val.
CSer	49	3.22	47.74	0.000	100	6.57	7.19	0.007	141	9.26	0.95	0.329
CDis	55	3.61	62.77	0.000	109	7.16	13.24	0.000	161	10.57	0.54	0.461
CSta	103	6.76	223.43	0.000	180	11.82	109.63	0.000	241	15.82	49.68	0.000
Ener	24	1.57	4.34	0.037	54	3.55	7.52	0.006	89	5.84	33.85	0.000
Fina	38	2.49	24.29	0.000	90	5.91	2.51	0.113	137	9.00	1.76	0.184
Heal	62	4.07	82.01	0.000	123	8.08	25.79	0.000	175	11.49	3.60	0.058
Indu	55	3.61	62.77	0.000	100	6.57	7.19	0.007	161	10.57	0.54	0.461
Mate	47	3.08	43.06	0.000	98	6.43	6.08	0.014	150	9.85	0.04	0.844
Real	52	3.41	55.07	0.000	119	7.81	21.83	0.000	178	11.69	4.60	0.032
Tech	50	3.28	50.14	0.000	92	6.04	3.27	0.071	140	9.19	1.13	0.288
Util	67	4.40	96.77	0.000	124	8.14	26.82	0.000	176	11.56	3.92	0.048

The Kupiec test only looks at how many times the actual loss goes beyond corresponding VaR. It doesn't consider the sequence of the violations which signifies a problem with the independence of violations. The accurate VaR measure should handle both unconditional coverage and independence (Campbell, 2005). One such test of conditional coverage which tests for both proportion of exceptions (unconditional coverage) and the clustering of exceptions (independence property) is the Christoffersen's Interval Forecast test, proposed by (Christoffersen, 1998). The method applies the Likelihood statistic testing framework as Kupiec, but extends the Kupiec test to jointly test for both unconditional coverage and independence. By combining the Kupiec test with independence, a joint test that examines both properties of an accurate VaR model is produced.

If we define n_{ij} as the number of observations i followed by j $(i,j = 0,1)$ where 1 indicates a violation and 0 indicates a nonviolation, then the test statistic

$$C = 2\,ln\left(\frac{(1-\pi_{01})^{n_{00}}\,\pi_{01}^{n_{01}}\,(1-\pi_{11})^{n_{10}}\,\pi_{11}^{n_{11}}}{(1-\alpha)^{N-F}\,\alpha^{F}}\right) \quad (7)$$

where $\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$ follows the chi-square distribution with two degrees of freedom (χ_2^2). If violations are independent, then the numerator and denominator will be approximately the same and the test statistic will be close to 0. Using this notation, the null hypothesis of independence claims states at t and t -1 are independent which means the violations at period t are preceded by a violation and by a nonviolation equally, i.e. $\pi_{01} = \pi_{11}$.

We present the results of the Christoffersen Test in Table 4. The null hypothesis of independence in addition to Kupiec Test's unconditional coverage asks for a significantly better performance of the VaR model. That is why the GARCH(1,1) model fails to satisfy both requirements of the Christoffersen Test substantially. Based on Table 4 we reject the joint null hypotheses of unconditional coverage and independence significantly. Note that this joint null is rejected for all sectors while testing for VaR(0.99), and VaR(0.95). The test results are similar with VaR(0.90) where the highest p-value is less than 1% for all sectors but Health Services with p-value 3.4%.

Table 4: Christoffersen Test Results

	VaR(0.99)		VaR(0.95)		VaR(0.90)	
	Test Stat	p-value	Test Stat	p-value	Test Stat	p-value
Cser	55.86	0.000	16.85	0.000	10.49	0.005
Cdis	71.44	0.000	23.25	0.000	12.93	0.002
Csta	228.52	0.000	113.20	0.000	54.91	0.000
Ener	17.17	0.000	16.60	0.000	48.60	0.000
Fina	42.34	0.000	19.40	0.000	13.16	0.001
Heal	90.65	0.000	34.19	0.000	6.73	0.034
Indu	77.98	0.000	21.29	0.000	13.16	0.001
Mate	55.49	0.000	16.56	0.000	14.16	0.001
Real	62.05	0.000	31.57	0.000	17.37	0.000
Tech	64.61	0.000	24.68	0.000	16.48	0.000
Util	101.53	0.000	42.47	0.000	16.08	0.000

Table 4 reveals that almost no test statistic for all VaR(0.99), VaR(0.95), and VaR(0.90) is less than 1%, 5%, and 10% χ_2^2 right tail areas; 9.210, 5.991, and 4.605, respectively. The only exception is VaR(0.90) for Health Care which is 6.93 marking a p-value of 3.4%.

CONCLUDING REMARKS:

This piece of research has two primary sources of motivation: performance of the GARCH(1,1) model in handling risk and the risk aspects of US sectors around the pandemic. The pandemic was so sudden and it had unprecedented impact on financial markets to present an open lab for testing GARCH model's capability to compute VaR.

The pandemic definitely exerted massive pressure on all US sector indexes. All such indexes plummeted by the beginning of the pandemic and the volatility of them increased sharply. On the other hand, the response of US sector indexes to the pandemic were idiosyncratic. Energy, for instance, followed a totally different path from others.

Essential descriptive statistics reveal that the Energy sector has the highest volatility. Based on VaR(0.90) of Kupiec Test we could not find enough evidence to reject the null hypothesis and this test reports the highest p-values from Materials, Consumer

Discretionary/Consumer Cyclical, Industrials, Communication Services and Information Technology.

Finally, Christoffersen Test results announce that GARCH(1,1) is not able to handle both unconditional coverage and independence simultaneously.

References (Bibliography)

Ahadiat, A., Kesumah, F.S.D. (2021). Risk Measurement and Stock Prices during the COVID-19 Pandemic: An Empirical Study of State-Owned Banks in Indonesia. Journal of Asian Finance, Economics and Business, 8(6), 0819–0828.

Baker, S.R., Bloom, N., Davis, S.J., Kost, K., Sammon, M., Viratyosin, T. (2020). The Unprecedented Stock Market Reaction to COVID-19. The Review of Asset Pricing Studies, 10(4), 742–758.

Berkowitz, J., Christoffersen, P., Pelletier, D. (2011). Evaluating Value-at-Risk Models with Desk-Level Data. Management Science, 57(12), 2213-2227.

Kupiec, P. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives, 2, 174–184.

Campbell, S. D. A Review of Backtesting and Backtesting Procedure. Finance and Economics Discussion Series, 2005-21. Board of Governors of the Federal Reserve System (U.S.). [Online]. Available on: http://www.federalreserve.gov/

Christoffersen, P.F., (1998). Evaluating Interval Forecasts, International Economic Review, 39(4) , 841-862

Christoffersen, P.F., Hahn, J., Inoue, A. (2001). Testing and Comparing Value-at-Risk Measures. Journal of Empirical Finance, 8(3), 325–342.

Hansen, B. E. (2011). Threshold Autoregression in Economics. Statistics and Its Interface, 4(2), 123-127.

Jorion, P. (2007). Value at Risk: The New Benchmark for Managing Financial Risk. The McGraw-Hill Companies, Inc.

McAleer, M., & Da Veiga, B. (2008). Single-Index and Portfolio Models for Forecasting Value-at-Risk Thresholds. Journal of Forecasting, 27(3), 217-235.

Omari, C, Mundia, S, Ngina, I, (2020). Forecasting Value-at-Risk of Financial Markets under the Global Pandemic of COVID-19 Using Conditional Extreme Value Theory. 10, 569-597

Orhan, M. and Koksal, B. (2012). A Comparison of GARCH Models for VaR Estimation. Expert Systems with Applications, 39(3), 3582–3592.

Shaik, M., Padmakumari, M, (2022). Value-at-Risk (VAR) Estimation and Backtesting During COVID-19: Empirical Analysis Based on BRICS and US Stock Markets. Investment Management and Financial Innovations, 19(1), 51-63.

DECISION SCIENCES INSTITUTE
Supply Chain Management in the AACSBS Accredited Business Schools Core

Md. Ziad Haidar
Eastern Illinois University
ziadhaidar05@gmail.com

Dr. Gurkan Akalin
The University of Virginia's College at Wise
ygd9my@uvawise.edu

Dr. Abdou Illia
Eastern Illinois University
aillia@eiu.edu

ABSTRACT
Supply chain management received significant attention during the covid-19 pandemic and during the war in Ukraine. We examine if the AACSB accredited business schools include SCM as part of their core curriculum.

KEYWORDS: Supply Chain Management, Core Curriculum, AACSB, Operations Management, Business Analytics

INTRODUCTION

Global and domestic supply chain crisis became a constant theme for 2019 and beyond starting with the Covid-19 pandemic and continue with the ongoing Russian invasion of Ukraine. Steady supply of many items all the way from baby formula to car-chips became increasingly complicated and at increasingly higher costs. With these much ongoing difficulties for businesses, supply chain management education should be at the top for business schools to be included in their core business programs. In this study, we examine the AACSB (Association to Advance Collegiate Schools of Business) schools for their supply chain management coverage. We do this by examining all the core classes of these schools. AACSB is the most accepted and desirable accreditation for the business schools (Jalbert, T., et al., 2011). Today they cover 542 schools (mid 2022) across all states including business schools with Ph.D. programs, Master's programs and undergraduate programs (AACSB, 2022). These accredited schools are both private and public schools which covers roughly 84% percentage of all business schools in the US. In terms of non-AACSB accredited schools, there are various reasons why they are not accredited but this article focuses only the ones that are AACSB accredited schools. The reason to focus on only AACSB schools is accreditation process ensures at least some guidance and "quality check" yet they cover significant percentage of all business schools in the US (Brink, K., & Smith, C., 2015). We only focus on the schools that are in the US to make sure our study is comprehensive yet manageable with the number of schools we are working on. We will not try to make a generalization for the global business education; we are only focusing on the US.

Business education is evolving similar to many other academic areas; however, the need to update what business schools teach changes rapidly (Tarabasz, A., et al., 2018) and, in many cases, faster than other fields. A good example is the subject of globalization which was assumed to be the dominant force in global trade up until Covid-19 pandemic hit and the war in Ukraine started. Now, there is the discussion of de-globalization (García-Herrero, A. 2019; Kornprobst, M., & Paul, T. V. 2021).This incidence shows that it is necessary to review the curriculum of business education to better align the needs of upcoming years (Bratianu, C., et al., 2021). Almost all business schools include classes in the format of business foundations, business core and business electives in their undergraduate degrees. Master's degrees also follow the same structure for many programs, but it is easier to find degrees with reduced number of core classes which allows more flexibility in their electives. Additionally, Master's degrees do not have to be comprehensive in subject breath to cover most of the business needs. A master's degree in taxation can be offered by a business school but does cover common subjects such as operations management or statistics. MBA degrees tend be more comprehensive along with undergraduate business degrees. In this study, we focus ourselves into the undergraduate business degrees which follow the strictest guidelines in terms of subject coverage. MBA degrees can be comprehensive but can be tailored into executive needs of students. MBA degrees may aim to cover the breath of business subjects through requiring prerequisites or exams. They are harder to quantify to assess what business schools value most in their teachings (Bennis, W. G., & O'Toole, J., 2005). AACSB accreditation does not require a specific set of classes for the undergraduate business degrees (Palocsay, S. W., & Markham, I. S. 2014); however, each business school needs to justify their curriculum process, they need to have peer schools which they compare themselves (Trapnell, J. E., 2007).

Research Question

In our study we aim to address following research question:

- What is the coverage of Supply Chain Management (SCM) in business core in business undergraduate degrees?

This study is important for evaluating SCM coverage of the education for the next generation of business professionals across the US; it is timely considering all the significant SCM challenges across the world.

Research Agenda

In this study, we examine all AACSB accredited 542 business schools who offer a business undergraduate degree. In other words, we are excluding Master's or Ph.D. granting institutions without a business undergraduate degree. A business undergraduate degree is any degree that is approved to use AACSB business accreditation. An accounting degree and marketing degree can be a business degree as long as they are approved by the AACSB accreditation. We will examine the business core requirements along with their prerequisites of each business school to see if there is any course with significant SCM coverage. We then compare this with the results from 2015-2016 of Akalin, G et al. (2016) paper to see the changes in the business core across the schools. Lastly, we do a survey to understand what the reasons schools do not include are SCM in their core program.

LITERATURE REVIEW

Supply chain education in universities plays a crucial role in preparing students for the challenges and opportunities presented by complex and global supply chains. This literature review aims to examine the existing research on supply chain education in universities, focusing on the key themes, approaches, and recommendations put forward by scholars in the field.

Choudhury and Paul emphasize the need for a cross-disciplinary approach to supply chain education, given the substantial differences between service supply chains and traditional supply chains (Choudhury & Paul, 2019). In a systematic literature review, Chong, Chong, and Li explore the capabilities and performance outcomes of supply chain education in universities. They emphasize the significance of understanding and developing key competencies such as analytical skills, problem-solving abilities, and adaptability (Chong, W. K., Chong, T., & Li, D. , 2020). The review highlights the positive relationship between supply chain education and organizational performance, suggesting that well-educated professionals contribute to improved supply chain operations and overall business outcomes (Chong, W. K., Chong, T., & Li, D. , 2020). To bridge the gap between theoretical knowledge and practical application, universities need to design supply chain curricula that are aligned with industry requirements. Collaboration with industry partners through internships, projects, and case studies provides students with valuable experiential learning opportunities Supply chain curricula should include courses and training on industry-relevant technical skills. These may include data analytics, supply chain software systems, inventory management tools, and transportation optimization. Equipping students with these skills enhances their employability and prepares them for the technological advancements in supply chain management. (Braglia, M., & Frosolini, M. ,2008).

Supply chain education in universities is a dynamic field that continues to evolve to meet the demands of the industry. It is important of a cross-disciplinary approach, industry collaboration, curriculum design, technological integration, a global perspective, and sustainability considerations in supply chain education. By focusing on these areas, universities can equip students with the knowledge, skills, and competencies necessary to excel in the complex world of supply chain management.

What is SCM (Supply Chain Management)?

Supply chain management is the systematic and strategic coordination of the business functions, strategic tactics across these business functions within and across the businesses within the supply chain, which works with the purpose of improving the sustainable performance of the company itself as well as the entire supply chain (Mentzer et al., 2001). According to (Hugos, M. H., 2018), supply chain management coordinates the production, inventory, location and transportation among all the participants of the entire supply chain with the aim of achieving the ideal mix of responsiveness and efficiency, that will be the most effective for the specific market in which the company is operating. The Council of Supply Chain Management Professionals (CSCMP) (2013) defines SCM as follows. Supply Chain Management encompasses the planning and management of all activities involved in sourcing and procurement, conversion, and all logistics management activities. Importantly, it also includes coordination and collaboration with channel partners, which can be suppliers, intermediaries, third-party service providers, and customers. In essence, supply chain management integrates supply and demand management within and across companies. It includes all of the logistics management activities noted above, as well as manufacturing operations, and it drives coordination of processes and activities with and across marketing, sales, product design, finance, and information technology.

Why SCM should be a salient part of business curriculum

In today's volatile business world, everything is constantly changing, so is the field of supply chain management and this has enlightened the growing interest in supply chain resilience (Pettit, T. J., et al.,2019). Unforeseen events like diseases, political turmoil, earthquake, fuel crises and terrorism have raised awareness among academics and practitioners to give more emphasis on resilient supply chains as a tool for minimising the damaging impact (Tukamuhabwa, B. R., et al., 2015). The increase of uncertainties and disruptive events has motivated the business leaders to pay special attention to resilient supply chain management as a strategic tool (Fiksel 2015). Although supply chains throughout the world have always been disrupted by unforeseen events, lately they have been significantly affected by the much-reaching devastating COVID-19 outbreak, with catastrophic consequences (Boccaletti et al., 2020). According to a Fortune (2020) survey released on 21st February 2020, 94% of the businesses included on the Fortune 1000 list experienced supply chain disruptions owing to COVID-19 (Fortune, 2020). Such disruptions has led to a growing interest in the area of supply chain management as well as enhanced research interest in supply chain risk (Sodhi, M. S., et al., 2012). Moreover, the fourth industrial revolution is reshaping future education, opening our vision, and making us consider the knowledge and skills necessary to possess (Li, L., 2020).

SCM in the curriculum

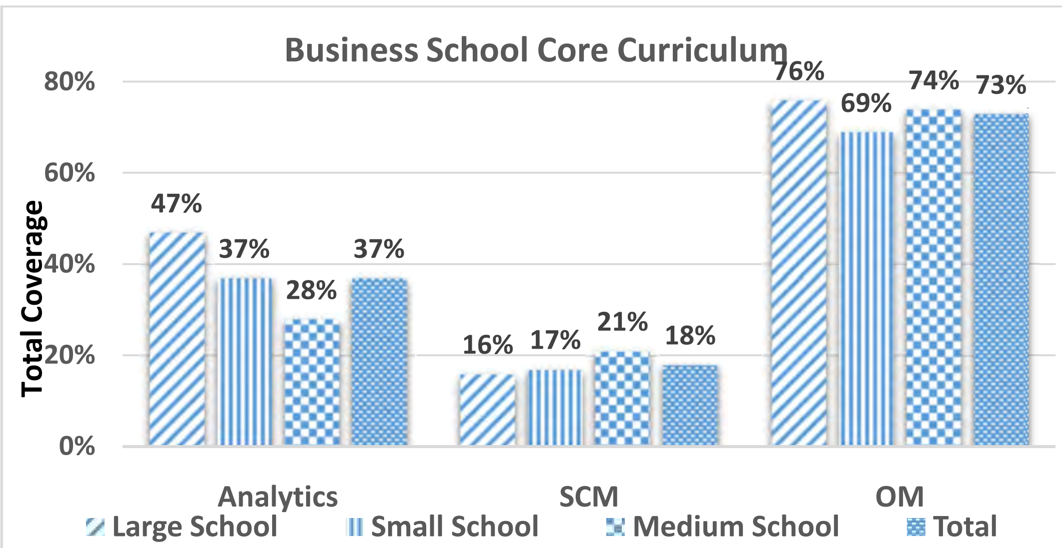
It is particularly important that educational institutions and the related industry work together to produce business graduates with the right set of knowledge and skills to be able to contribute to the current business world. For this reason, supply chain positions require not only well-trained graduates, but also those having broad range of skills and expertise (Jordan, C., &Bak, O., 2016). Ballou (2007) highlighted the need for greater understanding of the boundary spanning characteristic of the supply chain and its incorporation into the curriculum. Bak and Boloucher-Passet (2013), Hoffman (2005) acknowledged that the increased recognition of supply chain skill needs has caused the drive for a deeper understanding of supply chain in higher education.

Unfortunately, currently there is a dearth of coverage of supply chain disruption and risk management in the major supply chain management textbooks as well as in the related research publications (Ferguson, M. E., & Drake, M. J., 2021). Lancioni et al. (2001) indicated the lack of contribution from the education industry, especially in the undergraduate curriculum.(Sodhi, Son and Tang 2008) study reviewed the syllabi of MBA core courses in operations management and supply chain electives of the top-50 US business schools and found the topic of supply chain resilience management rarely covered. Akalin et al. (2016) found that only 2.91% of the AACSB accredited business schools offered supply chain management in their core curriculum

DATA

We examined all the AACSB accredited business schools' core curriculum. Core curriculum here refers to minimum required set of classes all business students are required to take irrespective of their majors. When the required set of classes includes options even if they are classified as the core classes, we did not include those options. The exception of this rule happens when the options are very close to each other in terms of content.

All the AACSB accredited business schools, irrespective of the size, have a curriculum largely focusing on operations management (OM) in their core courses, supply chain management (SCM) getting the least of the focus. This is more evident in the large schools than the small and medium schools. According to the data, 76% of the large schools have operations management (OM) in their supply chain curriculum, small and medium schools having 69% and 74% coverage, respectively.



Only 16% of the large schools have the core course supply chain management (SCM) in their curriculum, small and medium schools having only 17% and 21% respectively. Based on the core course supply chain management (SCM), large schools have the least emphasis on the core supply chain management course. On an average, 73% of the AACSB accredited schools have operations management (OM) in their supply chain curriculum, 37% of the schools have supply chain analytics but merely 18% of the business schools have the core course, supply chain management (SCM) in their supply chain curriculum. In this interesting to notice that relatively recent addition of business analytics has much more acceptance in the business core compared to the SCM being in the core. This calls for immediate attention on revisiting the existing supply chain curriculum and formalizing a curriculum that gives utmost attention to the c

CONCLUSION

This paper provides insight on the present status of supply chain curriculum in the AACSB accredited business schools. Clearly, Supply Chain Management (SCM) is yet to become the core business course in the US. With the data suggesting majority of the AACSB accredited schools having least focus on the core course Supply Chain Management, attention should be given to implement SCM as the core supply chain course. Comparing it with the Akalin, G., et al., (2016) study, where it was found that "6.04% of business schools have a core course that combines OM and SCM, while 2.91% of business schools offer only SCM in their core curriculum and less than 1% of business schools offer both OM and SCM in their core curriculum", current state of supply chain curriculum has improved ever since 2015-2016. However, comparing among operations management, supply chain management and analytics courses, the core course supply chain management has not been able to get the prime attention. Given the current state of the supply chain management core course in the AACSB accredited schools, we suggest that a revision of supply chain curriculum as necessary.

References:

Akalin, G., Huang, Z., & Willems, J. (2016). Is Supply Chain Management Replacing Operations Management in the Business Core Curriculum? *Operations and Supply Chain Management: An International Journal*, *9*(2), 119-130.

Bak, O., &Boulocher-Passet, V. (2013). Connecting industry and supply chain management education: exploring challenges faced in a SCM consultancy module. Supply Chain Management: An International Journal, 18(4), 468-479.

Ballou, R. H. (2007). The evolution and future of logistics and supply chain management. *European business review*.

Boccaletti, S., Ditto, W., Mindlin, G., &Atangana, A. (2020). Modeling and forecasting of epidemic spreading: The case of Covid-19 and beyond. *Chaos, solitons, and fractals*, *135*, 109794.

Braglia, M., & Frosolini, M. (2008). Supply chain management in industrial networks: contributions and perspectives. International Journal of Logistics Systems and Management, 4(6), 643-659.

Bratianu, C., Stanescu, D. F., &Mocanu, R. (2021). Exploring the knowledge management impact on business education. *Sustainability*, *13*(4), 2313.

Brink, K., & Smith, C. (2012). A comparison of AACSB, ACBSP, and IACBE accredited US business programs: An institutional resource perspective. *Business Education & Accreditation*, *4*(2), 1-15.

Chong, W. K., Chong, T., & Li, D. (2020). A systematic literature review of the capabilities and performance outcomes of supply chain education in universities. International Journal of Production Research, 58(3), 785-805

Choudhury, T. T., & Paul, S. K. (2019). A systematic literature review on the service supply chain: research trends and future directions. International Journal of Logistics Systems and Management, 32(2), 185-220.

Ferguson, M. E., & Drake, M. J. (2021). Teaching supply chain risk management in the COVID-19 Age: A review and classroom exercise. *Decision Sciences Journal of Innovative Education*, *19*(1), 5-14.

Fiksel, J. (2015). From risk to resilience. In *Resilient by design* (pp. 19-34). Island Press, Washington, DC.

Fortune. (2020). 94% of the Fortune 1000 are seeing coronavirus supply chain disruptions: Report. Retrieved March 30, 2020, from https://fortune.com/2020/02/21/fortune- 1000-coronavirus-china-supply-chain-impact/.

Hoffman, A. J., &Tessendorf, R. E. (2005, February). Artificial Intelligence based Fraud Agent to Identify Supply Chain Irregularities. In *Artificial Intelligence and Applications* (pp. 743-750).

Hugos, M. H. (2018). *Essentials of supply chain management*. John Wiley & Sons.

Jalbert, T., Jalbert, M., &Furumo, K. (2011). Does AACSB accreditation matter? Evidence from large firm CEOs. *Journal of Applied Business Research*, *27*(3), 93-106.

Jordan, C., &Bak, O. (2016). The growing scale and scope of the supply chain: a reflection on supply chain graduate skills. *Supply Chain Management: An International Journal*.

Lancioni, R., Forman, H., & Smith, M. F. (2001). Logistics and supply chain education: roadblocks and challenges. *International Journal of Physical Distribution & Logistics Management*.

Li, L. (2020). Education supply chain in the era of Industry 4.0. *Systems Research and Behavioral Science*, *37*(4), 579-592.

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business logistics*, *22*(2), 1-25.

Palocsay, S. W., & Markham, I. S. (2014). Management science in US AACSB international-accredited core undergraduate business school curricula. *Journal of Education for Business*, *89*(2), 110-117.

Pettit, T. J., Croxton, K. L., &Fiksel, J. (2019). The evolution of resilience in supply chain management: a retrospective on ensuring supply chain resilience. *Journal of Business Logistics*, *40*(1), 56-65.

Sodhi, M. S., Son, B. G., & Tang, C. S. (2012). Researchers' perspectives on supply chain risk management. *Production and operations management*, *21*(1), 1-13.

Sodhi, M. S., Son, B. G., & Tang, C. S. (2008). ASP, the art and science of practice: what employers demand from applicants for MBA-level supply chain jobs and the coverage of supply chain topics in MBA courses. *Interfaces*, *38*(6), 469-484.

Tarabasz, A., Selaković, M., & Abraham, C. (2018). The classroom of the future: Disrupting the concept of contemporary business education. *Entrepreneurial Business and Economics Review*, *6*(4), 231

Trapnell, J. E. (2007). AACSB International accreditation: The value proposition and a look to the future. *Journal of Management Development*.

Tukamuhabwa, B. R., Stevenson, M., Busby, J., &Zorzini, M. (2015). Supply chain resilience: definition, review and theoretical foundations for further study. *International Journal of Production Research*, *53*(18), 5592-5623.

DECISION SCIENCES INSTITUTE

Social Media Adoption and Use for Disaster Management: Exploring the Challenges for the Underserved Communities

Louis Ngamassi
Prairie View A&M University
longamassi@pvamu.edu

Modupe Ojumu
Prairie View A&M University
mfojumu@PVAMU.EDU

Munir Quddus
Prairie View A&M University
muquddus@pvamu.edu

ABSTRACT

Research shows that not all people are equally prepared to using social media for disaster management. In particular, the underserved communities, usually disproportionately impacted by disasters, are in a weak position to efficiently benefiting from social media use during disasters. Using a two-step research design, (i) we conduct a literature review of the challenges faced by the people in these communities in their use of social media for disaster management; (ii) we analyze data collected through interviews of key informants; and (iii) we compare and contrast the results from our analysis to the impeding factors already identified in the literature.

KEYWORDS: social media challenges, underserved communities, social media use, disaster management

INTRODUCTION

In recent years, policymakers and the research community have developed an interest in the potential benefits of social media especially in the fields of emergency and disaster management. Researchers have spent considerable amount of time exploring the adoption and use of social media for effective disaster management and collective decision-making (see Reuter et al., 2018 for a review). Social media is seen as a powerful tool that enables instant, low-cost, communication through common networks systems (Crowe, 2012; Neely and Collins, 2018). It also creates opportunities for multi-way dialogue (audio and video) resulting in improved flow of information and support (Bortree and Seltzer, 2009; Neely and Collins, 2018). Further, advances in mobile devices and their rapid spread in the society, have allowed this tool to become available and accessible to any user of the Internet. Furthermore, during disasters, social media can be used to organize communication, to strengthen the flow of information, and still be flexible to the changing disaster response needs (Sutton et al., 2008; Ngamassi et al., 2016; Ngamassi et al., 2016b). These attributes of social media make it an important tool that can be of tremendous help to the victims during the full cycle of disaster management.

Despite the fact that many studies have focused on the adoption and use of social media for disaster management and relief operations (e.g. Reuter et al., 2018; Karanasios et al., 2019; Karaye et al., 2019; Ngamassi et al., 2016; Ngamassi et al. 2016b; Ngamassi et al., 2017; Ngamassi et al., 2022; Ramakrishnan et al., 2019), a gap still exists with regards to explaining the reasons for the lack of participation by vulnerable communities in social media platforms. Understanding the challenges of adoption and use of social media platforms for disaster management and relief operations in these communities may be the first step to address the problem. Thus, we formulate our main research question as follows: What are the hurdles and challenges faced by members of the underserved communities to adopt and use social media for disaster related information management and sharing during disasters?

We address this question trough a two-step qualitative research approach that includes a scoping literature review (Pare et al., 2015) and semi-structured interview with community leaders and victims of disasters. Combining these two methods allows us to compare and contrast the challenges to the adoption and use of social media as perceived by people in the underserved communities and the barriers identified in the literature. The paper uses a four-component framework including, technology, organization, environment and people -TOEP- (DePietro et al., 1990; Baker, 2012) to analyze the challenges identified.
In the next section, we provide some background information on (i) the underserved communities and (ii) the TOEP framework. After that, we present the research methodology we used to collect and analyze data followed by a section on our data analysis. In the subsequent sections, we discuss our findings including the implications for research and practice, we identify the limitations of the study and suggest directions for future research.

BACKGROUND INFORMATION

Undeserved Communities

The underserved communities are defined as "a group or people whose circumstances create challenges to understanding or seeking information or to their ability to respond in the same way as the general population" (Karanasios et al. 2019, p. 2995). People in the underserved communities share a few specific characteristics such as a high percentage of senior citizens, isolated people, lower socio-economic people, gender, racial minorities, and marginalized people (Mannan et al., 2012; Karaye et al., 2019). These characteristics makes individuals and families in the underserved communities more susceptible to exclusion from social and financial relief services, unable to self-sustain on their own, and in need of extra assistance, particularly during distress and crises (Herrera and Gjøsæter, 2022).

Previous research suggests that underserved communities tend to suffer excessively from disasters as compared to other communities (Deacon, 2018; Karaye et al., 2019; Ramakrishnan et al., 2019; Bodenreider et al., 2019; Burger et al., 2019; Gan et al., 2021; Nejat et al., 2022). For example, findings from Klinenberg (2013) suggest that during the 2012 Hurricane Sandy and the 2013 Winter Storm Nemo, African-American communities were disproportionately impacted. Similarly, Squires (2017) reports that approximately forty six percent (46%) of the most severely impacted neighborhoods by the 2005 Hurricane Katrina were in the African-American community. Furthermore, Burger et al. (2019) found that economically vulnerable people were at greater risk and were disproportionally harmed by the 2012 hurricane Sandy. Research on the impacts of the Hurricane Harvey that struck Houston in November 2017 shows that communities of color and low-income communities suffered more extensive flooding than other communities; these communities were also found to be disproportionately exposed to toxic pollutants and environmental hazards (Bodenreider et al. 2019; Chakraborty et al., 2019; Collins et al., 2019). Another example is from a report by the Center for Disease Control and Prevention (2020) which estimated that African Americans are five times the risk of catching COVID-19 as compared to non-African American populations (CDC 2020).

TOEP framework

Researchers have devoted considerable time and energies investigating the adoption and use of information and communication technology. This literature suggests that most of the research are based on theoretical lenses and frameworks such as Resource-based Theory (Barney 1991), the Diffusion of Innovation (Rogers, 1995), Institutional Theory (Scott 1995), Technology Acceptance Model (Davis, 1989), and the Technology-Organization-Environment (DePietro et al., 1990).

The Technology-Organization-Environment (TOE) framework describes the major factors which influence the adoption and implementation of technological innovations at the organizational level (DePietro et al., 1990). The framework suggests that the adoption and implementation of technological innovations in an organization is largely influenced by three main factors including (i) technological context, (ii) organizational context, and (iii) environmental context. The technological context includes internal and external technologies relevant to the organization. The organizational context deals with the resources and the characteristics of the firm (e.g., scope, size, and managerial structure), and the environmental context refers to the external and internal environment of the firm such as the industry structure, firm's competitors, and regulatory environment. The TOE framework serves as a useful grouping of factors that facilitate or hinder the adoption of technological innovations (Haag and Eckhardt, 2014).

In the current study, we used an extended version of the TOE framework by adding people context as a fourth component. We call this the TOEP framework. The people context includes individual characteristics attitudes, perceptions and motivations with regards to innovation. Innovation adoption is largely dependent upon such individual characteristics of decision-actors (Awa et al., 2011; Awa et al., 2015; Awa et al., 2017). This extended version of the framework will provide a good basis for understanding the challenges faced by the undeserved communities when using social media for disaster management. For this study, the challenges considered within the framework were identified in the literature (e.g., Kuan and Chau, 2001; Martins and Oliveira, 2009; Oliveira and Martins, 2009; 2011). For instance, the technology context challenges include social media availability (lack of social media infrastructure), social medial accessibility (lack of technical expertise to access social media; affordability), social media complexity (technology readiness; technology integration; security applications), and information trustworthiness; the organization-related challenges include lack of social media strategy, lack of social media training program, lack of top management support to social media adoption and use; organization structure. The environment context includes digital divide, lack of technology, infrastructure failure, and government regulation. Finally, the people context challenges include the following: lack of education, low income, age, lack of IT skill and gender.

METHODOLOGY

In this paper, we use a two-step research approach to identify and analyze the challenges faced by people in the underserved or marginalized communities the access and use of social media for information management during the full cycle of a disaster. We follow the following research plan. First, we conducted a "scoping review" (Pare et al., 2015) of the literature related to these challenges. Second, we conducted a series of interviews (Mayring, 2015) and focus group discussions with key participants selected from the victims as well as responders during a major disaster in a rural county in south Texas (USA).

Scoping literature review

The purpose of a scoping literature review is to frame extant literature on a particular topic (Pare et al., 2015). Scoping literature reviews have been conducted and widely accepted in information systems research (e.g., Ngamassi et al., 2016; Tan et al., 2017). In this paper, we conducted a review of the literature in three steps.

First, we searched the ABI INFORM, ACM Digital library, EBSCOhost ,Science Direct and Scopus electronic databases for peer-reviewed papers from 2007 to 2022 to find social media challenges faced by the underserved communities that are identified in the literature. The focus was on the following three main research areas :(i) crisis informatics / information and communication technology / social media, (ii) disaster/crisis/emergency, and (iii) underserved communities/ minorities/ underrepresented. In our literature search, different keyword combinations were used: “social media,” crisis, emergency, disaster, underserved, minority and underrepresented. Our search query was written as follows: (“social media”) AND (disaster OR crisis OR emergency) AND (underserved OR minority* OR underrepresented). We present in Table 1 the concepts, terms, keywords and strings that we used for database search.

TABLE 1: Concepts and keywords used for database search

Concept	**Search Terms – Key Words, Strings**
Social Media	Social Media, Social Network, Social media Tool, Social Networking Site, Social Media Platform, , Twitter, Facebook, User-Generated Content, YouTube, Blog, Wiki
Disaster	Crisis, Disaster, Catastrophe, Emergency
Underserved Community	Underserved, minority, underrepresented, marginalized, low-income, African American

We limit the search to the abstract and keywords sections (as opposed to the full text) to allow us only to find papers that treat our search terms as an important part of their content. An initial examination of the titles and abstracts informed us that not all papers collected from the search were appropriate to our study. Subsequently, we excluded these papers from our sample. We present in Table 2, the criteria that we used to include and exclude papers.

TABLE 2: Criteria for Inclusion/Exclusion of Papers

Inclusion criteria
• Published research in the following areas: (i) social media, (ii) disaster/emergency/crisis, (iii) underserved communities; • Research published during the period 2007 -2022; • Published research that includes at least one of the following terms, “social media” and “disaster” or “crisis” or “emergency” and “underserved” or “minorities” or “underrepresented” in the keywords; • Published research in which the design and/or roles and/or uses and/or adoption and/or challenges of social media is discussed.

Second, in addition to the articles obtained from the database search, we also checked for relevant papers in the *International Journal of Disaster Risk Science* (IJDRS) and the *International Journal of Disaster Risk Reduction* (IJDRR), which provide an important venue for research on disaster risk reduction and information systems for crisis response and management. Further, we searched the proceedings (from 2010 to 2022) of *Information Systems for Crisis Response and Management Conference* (ISCRAM), a prominent conference on Information Systems (IS) with a track on "social media in crisis response and management."

Last but not the least, we used the ''snowball'' method to get additional papers. The snowball method involves a careful review of the reference list of the papers found in the first two steps and to look for additional papers pertinent to our study.

Overall, four papers were identified as relevant during the database search. The subsequent snowball search resulted in 9 additional papers, for a total of 14 relevant papers (Table 3)

TABLE 3: Results of Database Search

Database	Search results	Relevant papers
ABI INFORM	1	1
ACM Digital Library	3	0
AIS Electronic Library	0	0
EBSCOhost	13	0
IEEE Xplore	0	0
Science Direct	191	1
Scopus	67	3
Snowball	9	9
Total	**284**	**14**

Interviews

Through a series of -structured interviews (Mayring, 2015) and focus group discussions with key participants selected from disaster victims and disaster relief managers in a rural area of South Texas in the United States of America. We chose to conduct semi-structured interviews because they a allow interviewers to ask follow-up questions and interviewees to convey their experiences in a way that may not permitted by completely structured interviews. We guided each interview to cover specific topics but kept the process open for questions which arise during the interview process. The interview questions were designed to cover all four components of the TOEP framework, namely: Technology, Organization, Environment and People contexts. The interviews were conducted virtually. Among the ten key participants interviewed, there were five who represented disaster victims and five who were community leaders in charge of disaster management (one law enforcement officer, one county disaster management coordinator; one former Mayor, County Judge and a County disaster Manager). Each interview lasted between forty-five (45) to sixty (60) minutes. With permission from the participants, all the interviews were recorded and transcribed.

In addition to the interviews, we also conducted a focus group meeting with six participants, all disaster victims. The focus group was face-to-face lasting an hour and thirty minutes. The focus group meeting was recorded and transcribed. All the transcript data (interview and focus group) were uploaded to the NVIVO software, used to develop the coding structure and to codify the data.

Coding process

We used the "analytical induction" coding approach. This is a combination of deductive and inductive coding method (Epstein and Martin, 2004). For deductive coding, the four components of the TOEP framework, namely technology, organization, environment and people contexts were used. The inductive coding approach helped uncover information used to build sub-categories in the existing framework, or individual characteristics (e.g., age, gender, digital divide) that were not already identify in the components of the framework. During our iterative coding process, we reexamined the dataset multiple times to ensure a good understanding of the data. Moreover, the small size of the research team (three people) allowed for discussions leading to increased reliability of the coding process.

DATA ANALYSIS

Analysis of the literature review data

Our analysis of the data collected through the scoping literature review process shows that few researchers have explored the adoption and use of social media in disaster management by members of the underserved communities. The few papers have focused largely on the challenges relating to what we call the technological context. The question of information trustworthiness - rumors and deliberate false information circulating in social media during disasters is constantly identified across these papers as a big challenge. It was also often reported that people in the underserved communities are hesitant in using social media because of doubts regarding the quality and accuracy of information being shared. For instance, Morris et al., (2014) identifies three major challenges which prevent higher use of social media in reaching individuals with disabilities during emergencies and disasters. These include a) low use of social media, b) limited access to technology, and c) concerns regarding the accuracy and trustworthiness of social media data.

Although most papers in the literature have focused on challenges related to technology which hamper the adoption and use of social media in disaster management by underserved communities, a number of articles have also identified challenges which belong to the environment context. Digital divide and infrastructure failures were reported as among the biggest challenges. Haworth (2016) mentions the "digital divide" and limited access to the Internet due to lack of resources and infrastructure failure as major limitations for content-generated information sharing.

A number of papers in the literature have also looked into the challenges in the use of social media for managing disasters, which belong to what we call the "people context." These papers have identified the low-income levels, lack of IT skills, and old age as major impediments for the adoption and use of social media in disaster management by the members of the underserved communities (Morris et al., 2014; Haworth, 2016). It is interesting that this is similar to the concerns expressed in the literature concerned with diversity and inclusion issues.

Compared to the challenges in the three categories we discuss (technology, environment and people), challenges in the "organizational context" have received limited attention. A lack of access to social media and the existing organizational structure have been identified as major challenges in this literature. We present a concept matrix on our scoping review of the literature in Table 4 below.

TABLE 4: Concept Matrix on Scoping Literature Review

		Technology Context					**Organization Context**					**Environment Context**				**People Context**					
	Author(s), Year	SM Availability	SM	SM	SM	Information	SM	SM training	SM Access	Org Structures	Top	Digital divide	Technology	Infrastructure	Government	Education	Income	Age	IT Skill	Gender	Perceived
1	Keim and Noji (2011)												X								
2	Merchant et al. (2011)					X															
3	Alexander (2013)		X			X															
4	Morris et al. (2014)	X	X			X												X	X		
5	Anikeeva et al. (2015)											X		X							
6	Haworth (2016)											X		X							
7	Feldman et al.(2016)					X												X			
8	Majid and Spiro (2016)												X			X					
9	Neely & Collins (2018)									X						X		X			
10	Dragović et al. (2019)	X	X															X			
11	Leeson (2019)		X																		
12	Dargin et al. (2021)					X						X									
13	Ramakrishnan et al. (2022)											X									
14	Verma et al. (2022)					X															
	Total	2	4	0	0	6	0	0	0	1	0	5	2	2	0	2	0	4	1	0	0
		12					1					9				7					

Analysis of interview data

As mentioned earlier, we used the NVIVO software to "code" the data collected through interviewing key informants. In table 5 below, we provide a summary of the codebook.

TABLE 5: Challenges Codebook Generated from NVIVO Platform

Challenges	**% of Interviewees**	**# of References**
Environment context		25 (15.92%)
Digital divide	40%	6
Government regulation	20%	2
Infrastructure failure	60%	10
Technology support	30%	7
Organization context		18 (11.46%)
Lack of resources	30%	9
Organization structures	20%	2
Social media access	40%	7
People context		25 (15.92%)
Age	60%	8
Culture	20%	5
Education	20%	3
IT Skill	50%	9
Technology context		89 (56.70%)
Information overload	20%	2
Information trustworthiness	100%	26
Lack of contextual awareness	20%	7
Social media accessibility - barriers	50%	9
Social media availability	50%	19
Social media characteristics	50%	14
Social media complexity	40%	12

Similar to the findings suggested by the scoping literature review, the analysis of the interview data shows that members of the underserved communities consider challenges in the technology context to be most significant in their adoption and use of social media for information management during disaster. As shown in Table 3, more than half (56.70%) of the references are related challenges of this category. Specifically, participants in the study have expressed concern about the quality of information circulated through social media in times of crisis. All interviewees identified a lack of information trustworthiness as a major challenge. For instance, one interview participant, a community leader, said this:

> *"The disadvantage of it is that it is not always true. Things that are put on social media are at times not correct. I think social media sometimes if there are some individuals out there that try to cause fear and in these times they may post a picture or an image that looks bad." (participant#2)*

Another interview participant, a disaster victim, said:

> *"The thing I don't like about social media is that some things can get tied up, twisted words, and that's how people get fabricated lies and information through social media. It's not always true." (participant#10)*

In addition to the quality of information shared through social media during disaster, there was also a concern about the need for actionable information. An actionable information is formulated to easily provide answers to the Why, Who, Where, What, and When questions (Ngamassi et al., 2022). A community leader said:

> *"I really think the most important factor is not who is putting information out in time of crisis rather, is the information accurate and is it actionable" (participant#3)*

Social media availability (lack of social media infrastructure), social medial accessibility (lack of technical expertise to access social media; affordability), social media complexity (technology readiness; technology integration; security applications) were also identified as major technological related challenges to the adoption and use of social media for information sharing by the people of the undeserved communities during disasters. Roughly half of the interview participants mentioned either or all of these challenges. For example, one of the community leaders interviewed said:

> *"I think there are quite a few social media sites out there. I know they're evolving. I know they are changing. And that's one of the things that you know I think we need to work on." (participant#2)*

Sharing a personal experience, another community leader said:

> *"The things I dislike about it is that it provides a platform to anybody hiding behind their keyboard to attack someone else whether it's a personal attack or general criticism." (participant#4).*

As compared with challenges belonging to the technology context, participants to the interview perceived challenges related to the three other categories of the TOEP framework as of lesser magnitude. Approximately sixteen percent (15.92%) of the references mentioned equally in both the environmental and people categories, and less than twelve percent (11.46%) to the organization category. In line with the results revealed by the scoping review of the literature, digital divide and infrastructure failures were reported to be the biggest challenges in the environment context. Two of the community leaders interviewed reported respectively that:

> *"When this stuff goes down when you get bad reception, you are blind. …. You don't have access to power and your cell phone dies, you are finished." (participant#1).*

"So even that presents challenges. In some areas like Houston you know, Internet access is a given. It is taken for granted. Not everyone in our County has good Internet access. This is something we're still working on. We are working with different entities to try to make sure we get more broadband access to folks in the county which then will help us reach out to folks in the underserved communities and communicate better." (participant#4).

The lack of information technology (IT) skills and old age were reported as major challenges in the people context. For instance, one of the community leaders said:

"The other side about it, the technology side is that we realize that not everyone is well-verse with the Internet" (participant#2)

In the organization category, lack of resources was reported to be a big challenge.

"Like I stated earlier, our goal is to try to find our way on social media. At this time with limited staff, it is challenging for us to do that. If we want to do that, I think it will not be successful." (participant#2)

"Technology, especially if you're talking about the infrastructure, you have to have a very strong security infrastructure and communications infrastructure which we do not have." (participant#3)

DISCUSSIONS AND IMPLICATIONS

The motivation behind this research project was to examine the challenges faced by those in the underserved communities in the adoption and of social media to protect themselves during disasters with the goal to find ways to increase the use this new communication technology by all segments of the society during the full cycle of disaster management. Here we should recognize that increasing the use of social media has at least two dimensions – first, more people adopt and take advantage of social media especially in time of crises, and second, the quality of social media data improves. This would lead to greater and more balanced impact of social media as a tool for preparations and seeking relief from disasters and calamities.

As mentioned earlier, previous studies show the benefits of the use of social media for disaster management (Spiro et al., 2012; Semaan and Mark, 2012; Van Wyk and Starbird, 2020). Findings from previous studies also suggest that access and use of social media is mostly confined to the young, those with education, those with good English language skills and resources (e.g. ; Neely and Collins, 2018; Spielhofer et al., 2019). Research also indicates that underserved communities are generally affected more during disasters (Karaye et al., 2019; Deacon, 2018; Squires, 2017). Yet, there is a dearth of studies that examine the opportunities and/or challenges of the use of social media for disaster management by the people in the underserved communities.

Employing a broad scoping review of the literature complemented by interviews of key participants, we identify in the scanty existing literature a number of challenges faced by these communities. The combined results from these two approaches were captured using four categories provided by the TOEP framework. The challenges identified from both approaches were consistent in order of importance ranging from the technology context (most important), people context, environment context, and organization context (least important). Technology related challenges have also been identified cross literature as a significant impediment to the use of social media for disaster management irrespective of the community (e.g. Spielhofer et al., 2019). This may also be a consequence of the digital divide. Findings from previous research that investigated the digital divide in terms of access and affordability of technology, suggest that the accessibility and affordability for underserved communities is significantly lower than that for other communities (van Deursen and van Dijk 2014; Ram et al, 2022).

When asked why during disasters – nature or manmade - some groups (vulnerable populations) are deprived from the many benefits of social media in getting assistance and support, we realize that this issue has moral dimensions. It is well known that those in the racial minorities are overrepresented among the low-income and low-asset households, which are precisely the groups which fail to take full advantage of social media during periods of distress and disasters. In other words, we have a “diversity, inclusion and equity” dimension to this issue, one that the literature has failed to address adequately. If we wish that the benefits of social media should be widely available to all segments of the society for the “greater good,” of the society, scholars and policymakers must ensure that the low-income, racial minorities, and otherwise “vulnerable” households have equal access and sufficient training in the use of social media to protect their lives, property and livelihoods.

This study has both theoretical implications for the scientific community and practical implications for all the stakeholders involved in humanitarian disaster management and relief operations.

Implications for Theory

The theoretical implications of this study are twofold. As discussed earlier, there is a very scanty literature that examines the challenges of the adoption and use of social media for disaster management by the people in the underserved communities. This work contributes the increase that very limited literature. Second, this study extends the use of the technology-organization-environment-people (TOEP) framework in the field of disaster management. In this paper, the framework is used as a taxonomy to categorize in order to better understand the problems that the underserved communities faced when using social media for disaster management.

Implications for Practice

From a practical perspective, the findings from this research show that the major impediments to the adoption and use of social media for disaster management by the people of the undeserved communities are technology related. Thus, the stakeholders involved in disaster relief (e.g., government agencies, nongovernmental organizations) should put more effort into addressing these issues. In particular, information trustworthiness is seen both in the previous studies and by the key informants that were interviewed as the most severe challenge. Effort should be made to educate these people on how to identify fake news (e.g., reliable source of information). Effort should also be made in providing basic social media training to this community. To this end, outreach programs in collaboration with institutions of higher education (e.g., community colleges and four-year colleges) could be designed developed and implemented.

LIMITATIONS

Our work has limitations that should be acknowledged in the interpretation of the findings presented above. First, although for the scoping literature review, we have used five big databases (ABI INFORM, ACM Digital library, EBSCOhost, Science Direct and Scopus), our final results may have excluded relevant articles. For instance, some studies have found that Social Sciences journal articles are underrepresented in Scopus (Mongeon and Paul-Hus, 2016). The Second, limitation is that we conducted the search of database based on a set of pre-defined criteria. This means our final set of data has excluded book chapters, conference proceedings other than ISCRAM, and non-peer reviewed articles. However, our use of alternate keywords to examine titles and abstracts, and also the use of the ‘‘snowball’’ method to find additional relevant papers, we believe would have minimized the chances of an omitted paper of critical implications for our findings. Third, the study used a relatively small number of key informants. This may affect the generalizability of the findings. Such limitations can be reduced by increasing the number of participants drawn from a wider type of organizations. Fourth, the TOE framework is mostly used at the organizational level of analysis. In this study, although we used an extended version of the framework (adding the People context), we may not have efficiently captured the challenges belonging to the organizational context from our interview of key informants. All these limitations can offer starting point for future research directions.

This research points to a “gap” in the literature and investigates factors prevent the use of social media to its full potential by the underserved communities. We then mention a few factors responsible for this. One missing factor is the fact that a large segment of the population – especially those without a college degree – are “suspicious” of social media for privacy and other reasons. Many in this group are more susceptible to disinformation in the form of “conspiracy theories” since the Internet is rife with “misinformation” and “disinformation” which has convinced millions to shun a beneficial tool, since they feel one may be tracked by the state agencies. The fear of “big government” and “big business” is very real among an increasing number of citizens in this group. They may believe social media is another tool used to collect information on them and track their activities. There is no denying the fact that FB and some big companies have intentionally sold information on their users. Those who believe in conspiracies take this to the extreme, and thus shun technology including social media.

CONCLUSION

In this study, we investigated the challenges faced by the people in the underserved communities when adopting and using social media for disaster related information management and disaster relief operations. We used a two-step research method, scoping literature review and interview of key informants, to identify the challenges. We then used the TOEP framework as taxonomy to categorize and synthesize the results from both methods. Combining these two methods allowed us to compare and contrast the challenges related to the use of social media perceived by people in the underserved communities and the impeding factors already identified in the literature. Moreover, according to the triangulation method (Jick, 1979), a combination of different approaches to address a research question ensures more valid and comprehensive results.

Our study highlights a dearth of research that investigates the opportunities and/or challenges of the adoption and use of social media for disaster management by the people in the underserved communities and calls for more research in this area especially as the underserved communities suffer the most in times of crises. The findings from our research also suggest that the challenges belonging to the technology context of the TOEP framework were the one identified in both methods to be the greatest impediment for the adoption and use of social media by the people in the underserved communities for disaster management. Based on these findings, we made propositions to the stakeholders involved in disaster management to contribute to improving the effectiveness of the adoption and use of social media for disaster management and relief operations by the people in the underserved communities. Our propositions could also serve as directions for future research.

ACKNOWLEDGMENTS

This article was written based on work supported by the U.S. Department of Homeland Security under Grant Award Number, 2017-ST-062-000005.

REFERENCES

Alexander, D. 2013. Social Media in Disaster Risk Reduction and Crisis Management. Science and engineering ethics. 20. 10.1007/s11948-013-9502-z. Retrieved March 2, 2020.

Anikeeva, O., Steenkamp, M. and Arbon, P. 2015. "The Future of Social Media Use During Emergencies in Australia: Insights from the 2014 Australian and New Zealand Disaster and Emergency Management Conference Social Media Workshop." *The Australian Journal of Emergency Management* 30 (1), 22-26.

Awa, H. O., Ukoha, O., and Igwe, S. R. 2017. Revisiting technology-organization-environment (TOE) theory for enriched applicability. *The Bottom Line*.

Awa, H., Baridam, D. and Nwibere, B. 2015. Demographic determinants of e-commerce adoption: a twist by location factors, *Journal of Enterprise Information Management*, *28*(3), 325-346.

Awa, H., Eze, S., Urieto, J. and Inyang, B. 2011. “Upper echelon theory: a major determinant of information technology (IT) adoption by SMEs in Nigeria,” *Journal of Systems and Information Technology*, *13*(2), 144-162.

Baker, J. 2012. The technology–organization–environment framework. In Information systems theory (pp. 231-245). Springer.

Barney, J.B. 1991. Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99-120.

Bodenreider, C., Wright, L., Barr, O., Xu, K., and Wilson, S. 2019. Assessment of social, economic, and geographic vulnerability pre-and post-Hurricane Harvey in Houston, Texas. *Environmental Justice*, *12*(4), 182-193.

Bortree, D.S. and Seltzer, T. 2009. Dialogic strategies and outcomes: An analysis of environmental advocacy groups’ Facebook profiles. *Public Relations Review,* 35(3)*,* 317-319.

Burger, J., Gochfeld, M., and Lacy, C. 2019. Ethnic differences in risk: experiences, medical needs, and access to care after hurricane Sandy in New Jersey. *Journal of Toxicology and Environmental Health, Part A*, *82*(2), 128-141.

CDC 2020. COVID-19 in racial and ethnic minority groups. https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html

Crowe, A. 2012. *Disasters 2.0: The application of social media for modern emergency management*. Taylor and Francis Group Publisher.

Chakraborty, J., Collins, T. W., and Grineski, S. E. 2019. Exploring the environmental justice implications of Hurricane Harvey flooding in Greater Houston, Texas. *American journal of Public Health*, *109*(2), 244-250.

Collins, T. W., Grineski, S. E., Chakraborty, J., and Flores, A. B. 2019. Environmental injustice and Hurricane Harvey: A household-level study of socially disparate flood exposures in Greater Houston, Texas, USA. *Environmental Research*, *179*, 108772.

Davis, F. 1989. Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-40.

Davis, T. C., Arnold, C. L., Mills, G., and Miele, L. 2019. A qualitative study exploring barriers and facilitators of enrolling underrepresented populations in clinical trials and biobanking. *Frontiers in Cell and Developmental Biology*, *7*, 74.

Deacon, B. 2018. Emergency announcements alone during disasters not reliable in saving lives. ABC News. https://www.abc.net.au/ news/2018-01-06/emergency-announcements-alone-during-disasters-not-reliable/9304666.

DePietro, R., Wiarda, E., and Fleischer, M. 1990. ‘The context for change: organization, technology, and environment’, In L. G. Tornatzky and M. Fleischer (eds) *The processes of technological innovation*, Lexington Books, pp. 151-175.

Dragović, N., Vasiljević, Đ. and Stankov, U. 2019. Go social for your own safety! Review of social networks use on natural disasters – Case studies from worldwide. *Open Geosciences*, *11*(1), 352-366. doi:10.1515/geo-2019-0028.

Epstein, L., and Martin, A. 2004. Coding variables. In K. Kempf-Leonard (Ed.), The handbook of social measurement.

Feldman D., Contreras S., Karlin B., Basolo V., Matthew R., Sanders B., Houston D., Cheung W., Goodrich K., Reyes A., Serrano K., Schubert J., and Luke A. 2016. Communicating flood risk: Looking back and forward at traditional and social media outlets, *Int. Journal of Disaster Risk Reduction*, *15*43–51. doi:10.1016/j.ijdrr.2015.12.004.

Gan, C. C. R., Oktari, R. S., Nguyen, H. X., Yuan, L., Yu, X., Alisha, K. C., ... and Chu, C. 2021. A scoping review of climate-related disasters in China, Indonesia and Vietnam: Disasters, health impacts, vulnerable populations and adaptation measures. *International Journal of Disaster Risk Reduction*, *66*, 102608.

Gill, A. and Bunker, D. 2012. Crowd sourcing challenges assessment index for disaster management. In: *Proceedings of the Americas conference on information systems*. Seattle, WA, USA.

Goodchild, M. F. and Glennon, J. A. 2010. Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth 3*(3), 231-241.

Haag, S., and Eckhardt, A. (2014). Organizational cloud service adoption: A scientometric and content-based literature analysis. *Journal of Business Economics*, *84*(3), 407-440. doi:10.1007/s11573-014-0716-6

Haworth, B. 2016. Emergency management perspectives on volunteered geographic information: Opportunities, challenges and change. *Computers, Environment and Urban Systems* (57), 189- 198.

Herrera, L. C., and Gjøsæter, T. 2022. Community Segmentation and Inclusive Social Media Listening. Proceedings of the 19th *International Conference on Information Systems for Crisis Response and Management (ISCRAM),* Tarbes, France May 22-25.

Jick, T. D. 1979. Mixing qualitative and quantitative methods: Triangulation in action." *Administrative Science Quarterly 24*(4), 602-611.

Karanasios, S., Cooper, V., Balcell, M. P., and Hayes, P. 2019. Inter-organizational collaboration, information flows, and the use of social media during disasters: A focus on vulnerable communities.

Karaye, I. M., Thompson, C., and Horney, J. A. 2019. Evacuation shelter deficits for socially vulnerable Texas residents during Hurricane Harvey. *Health Services Research and Managerial Epidemiology*, *6*, 2333392819848885.

Keim, M. and Noji, E. 2011. Emergent Use of Social Media: A New Age of Opportunity for Disaster Resilience. American journal of disaster medicine. Vol 6. Pp. 47-54

Kuan, K.K.Y. and Chau, P.Y.K. 2001. A perception-based model for edi adoption in small businesses using a technology-organization-environment framework, *Information & Management*, *38*(8), 507-521.

Majid, A. M. and Spiro E. S. 2016. Crisis in a foreign language: Emergency services and limited English populations. *Ethical, legal, and social issues proceedings of the ISCRAM 2016 conference,* Rio de Janeiro, Brazil. http://idl.iscram.org/files/amirahmmajid/2016/1363_AmirahM.Majid+EmmaS.Spiro2016.pdf

Mannan, H., MacLachlan, M., McVeigh, J., and Consortium, E. 2012. Core concepts of human rights and inclusion of vulnerable groups in the United Nations Convention on the rights of persons with disabilities. *Alter*, *6*(3), 159-177.

Martins, M. and Oliveira, T. 2009. Determinants of e-commerce adoption by small firms in Portugal, *Proceedings of the 3rd European conference on information management and* *evaluation.* Gothenburg, Sweden, September 2009, 328-338.

Mayring, P. 2015. Qualitative content analysis: Theoretical background and procedures. In: Bikner-Ahsbahs, A., Knipping, C. and Presmeg, N. (eds) *Approaches to Qualitative Research in Mathematics Education*. Springer, Dordrecht, 365-380.

Merchant, R., Elmer, S. and Lurie, N. 2011. Integrating social media into emergency-Preparedness efforts. *The New England Journal of Medicine*. *365*(4), 289-91.doi:10.1056/NEJMp1103591.

Mongeon, P., Paul-Hus, A., 2016. The journal coverage of web of science and Scopus: a comparative analysis. *Scientometrics 106*(1), 213–228.

Morris, J. T., Mueller, J. L., and Jones, M. L. 2014. Use of social media during public emergencies by people with disabilities. *The Western Journal of Emergency Medicine*, *15*(5), 567–574. https://doi.org/10.5811/westjem.2014.4.21274

Neely, S. and Collins, M. 2018. Social media and crisis communications: A survey of local governments in Florida. *Journal of Homeland Security and Emergency Management*, *15*(1). doi:10.1515/jhsem-2016-0067.

Nejat, A., Solitare, L., Pettitt, E., and Mohsenian-Rad, H. 2022. Equitable community resilience: The case of Winter Storm Uri in Texas. *International Journal of Disaster Risk Reduction*, 103070.

Ngamassi, L., Ramakrishnan, T., and Rahman, S. 2016. Use of social media for disaster management: A prescriptive framework. *Journal of Organizational and End User Computing, 28*(3).

Ngamassi, L., Ramakrishnan, T., and Rahman, S. 2016b. Examining the role of social media in disaster management from an attribution theory perspective. *Proceedings of the 13th international conference on information systems for crisis response and management (ISCRAM),* Rio de Janeiro, Brazil, May 22-25, 2016.

Ngamassi, L., Malik, A., Zhang, J., and Edbert, D. 2017. Social media visual analytic toolkits for disaster management: A review of the literature. *International conference on information systems for crisis response and management (ISCRAM)*, Albi, France, May 21-24, 2017.

Ngamassi, L., Shahriari, H., Ramakrishnan, T., and Rahman, S. 2022. Text mining hurricane harvey tweet data: Lessons learned and policy recommendations. *International Journal of Disaster Risk Reduction*, *70*, 102753.

Oliveira, T. and Martins, M. F. 2009. Determinants of information technology adoption in Portugal, ICE-B 2009. *Proceedings of the international conference on e-business*, Milan. Italy, July, 264-270.

Oliveira, T., and Martins, M. F. 2011. Literature review of information technology adoption models at firm level. *Electronic Journal of Information Systems Evaluation*, *14*(1), 110-121.

Pare, G., Trudel, M. C., Jaana, M., and Kitsiou, S. 2015. Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management,* 52, 183-199.

Ramakrishnan, T., Ngamassi, L., & Rahman, S. (2022). Examining the factors that influence the use of social media for disaster management by underserved communities. *International Journal of Disaster Risk Science*, *13*(1), 52-65.

Ramakrishnan, T., Ngamassi, L. and Rahman, S. 2019. Social Media in Disaster Management: Exploring the Information Exchange Behavior in the Underserved Communities*, Proceedings of the 16th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*. Valencia, Spain May 23-25, 2019

Reuter, C., Hughes, A. L., and Kaufhold, M. A. 2018. Social media in crisis management: An evaluation and analysis of crisis informatics research. *International Journal of Human–Computer Interaction*, *34*(4), 280-294.

Rogers, E.M. 1995. *Diffusion of innovation,* 4th edition. The Free Press, New York.

Sawyer, S., and Huang, H. 2007. Conceptualizing information, technology, and people: Comparing information science and information systems literatures. *Journal of the American Society for Information Science and Technology*, *58*(10), 1436-1447.

Scott, W. R. 1995. *Institutions and organizations*. Thousand Oaks.

Spielhofer, T., Hahne, A., Reuter, C., Kaufhold, M., and Schmid, S. 2019. Social Media Use in Emergencies of Citizens in the United Kingdom. In Z. Franco, J. J. González, & J. H. Canós (Eds.), *Proceedings of the 16th International Conference on Information Systems for Crisis Response And Management*. Valencia, Spain: Iscram

Spiro, E. S., Sutton, J., Greczek, M., Fitzhugh, S., Piersk, N., and Butts, C. (2012, June). Rumoring during extreme events: A case study of Deepwater Horizon 2010. In *Proceedings of WebSci 2012*. June, Evanston, IL

Squires, G. D. (2017). Harvey is not a natural disaster. *The American Prospect*. https://prospect.org/article/harvey-not-natural-disaster

Sutton, J., Palen, L. and Shklovski, I. 2008. Backchannels on the front lines: Emergent uses of social media in the 2007 Southern California wildfires. In *Proceedings of the 5th International ISCRAM Conference*. *Seattle, Washington.*

Tan, M. L., Prasanna, R., Stock, K., Hudson-Doyle, E., Leonard, G., and Johnston, D. 2017. Mobile applications in crisis informatics literature: A systematic review. *International Journal of Disaster Risk Reduction*, *24*, 297-311.

van Deursen, A.J.A.M., and J.A.G.M. van Dijk. 2014. The digital divide shifts to differences in usage. New Media & Society 16(3): 507–526.

Van Wyk, H., and Starbird, K. 2020. Analyzing social media data to understand how disaster-affected individuals adapt to disaster-related telecommunications disruptions. *Proceedings of the 17th ISCRAM conference* – Blacksburg, VA, USA, May 2020. Amanda Lee Hughes, Fiona McNeill and Christopher Zobel, (Eds.). 704-717.

Yu, M., Yang, C., and Li, Y. 2018. Big data in natural disaster management: A review. *Geosciences*, *8*(5), 165.

DECISION SCIENCES INSTITUTE

Social Media Capabilities and Customer Engagement impact on Decision-Making Performance

Hussam Alansari
Earl. G. Graves School of Business & Management, Morgan State University, 4100 Hillen Road, Baltimore, Maryland 21251, huala6@morgan.edu

ABSTRACT

Social media capabilities are presenting opportunities for organizations to improve the quality of managerial decisions and customer engagement.

To examine the impact of social media capabilities, and customer engagement, on decision-making performance. This study will draw from media richness theory and media synchronicity theory. This study suggests media richness theory and media synchronicity theory for integrating several SMC and customer engagement dimensions, their cooperative effects on decision-making performance.

This study will contribute to the existing literature by developing empirical evidence that confirms the relationship and significance of SMC and CE in improving decision-making performance. Moreover, this study will instruct and draw the attention of managers on the importance of SMCs in decision-making performance thus in a competitive advantage for the sports' firms.

Key words: Social Media Capabilities (SMCs), Decision-Making Performance, Customer Engagement (CE)

INTRODUCTION

The potential of social media capabilities to provide firms competitive advantage has been a topic of interest to practitioners and academicians (Abeza, 2015; Al-Qaysi & Al-Emran,2017; Luckshay, 2019; and Wang, 2017).

Social media has transformed how people interact with one another and reshaping how businesses interact with their consumers. Companies are progressively investing in the establishment of social media platforms and methods to augment customer-firm interactions, with 75 percent of Internet users engaged in social media (Forrester Research 2008). Frequently, customers use social media to explicit opinions, feelings, and suggestions about the services they have been used. Customers interaction on social media sites help organizations to know their behaviors during certain event and can provide valued knowledge to help organizations to improve their decision-making quality. (He, 2016) developed data analytics approach which can help organizations to gain a competitive advantage to extract information about people opinions and believes which can identify what has happed and predict what may happen later.

Social media capabilities are presenting opportunities for organizations to improve the quality of managerial decisions and customer engagement (Singh, 2020; Kaplan & Haenlein, 2010). In the same context, number of studies have focused on how SMC has been used by customers to build community and promote preferred representations of athletes and sport organizations (Sanderson, 2013). In addition, SMC can offer opportunities for analyzing customers engagement behavior and enhance relational satisfaction with the public (Del Giudice, 2016). This behavior analysis can provide organizations with information about people opinions and concerns in certain event, which help organizations to develop new managerial decisions.

To examine the impact of social media capabilities, and customer engagement, on decision-making performance. this study will draw from media richness theory and media synchronicity theory.

In this context, the main purpose of this study focuses on analyzing the effect of social media capabilities and customers engagement on decision-making performance. This study will contribute to the existing literature by developing empirical evidence that confirms the relationship and significance of SMC and CE in improving decision-making performance. Moreover, this study will instruct and draw the attention of managers on the importance of SMCs in decision-making performance thus in a competitive advantage for the firms.

The outline of the paper is divided into three sections. The remaining of the paper is organized as follows. Section 2 is discussing the literature review. Section 3 is the research methodology.

LITERATURE REVIEW

The review of the literature formerly covers the three significant constructs of the study, namely social media capabilities, customer engagement, and decision-making performance.

Social Media Capabilities

Kaplan and Haenlein (2010) define social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of user-generated content." By April 2011, the Wikipedia definition was simplified as "social media are media for social interaction, using highly accessible and scalable communication techniques. Social media is the use of web-based and mobile technologies to turn communication into interactive dialogue." Examples of social media includes Collaboration projects (eg., Wikipedia), blogs and microblogs (e.g., Twitter), content communities (e.g., YouTube), social networking (e.g., Facebook), virtual gaming worlds, and virtual social worlds (He, 2016). Companies utilize social media to develop direct relationships with their consumers, enhance website traffic, uncover new business prospects, form communities, disseminate content, collect customer feedback, and generally promote their brand (Michaelidou et al., 2011)

(Luxton et al., 2015) investigated how social media capabilities impact decision-making framework. the study highlighted the following characteristics: (1) the ability to provide a clear decision-making series to direct overloaded and unsure managers faced with social media choices; (2) the ability to integrate traditional media and the complementary strengths of social media; (3) the able to discuss SM implementation challenges; and (4) the guarantee organizational structures, cultures, and processes facilitate social media integration and implementation.

Customer Engagement

A high number of social media postings by a customer indicates that the client is more engaged with the activity and the company, which raises the degree of engagement between the customer and the firm (Sashi, 2012). This process, in which customers express their sentiments of connection with the activity and the firm by uploading comments, photos, and videos, is extremely beneficial to businesses. The technique enables businesses to always communicate with customers, not only when they are really using the service (Lim, Hwang, Kim, & Biocca, 2015). The idea of engagement refers to a motivational condition in which an individual's (the engagement subject) primary interaction experiences with a certain item or agent (the engagement object) are emphasized (Hollebeek et al., 2014). The customer or consumer is the engagement subject, whereas the engagement object might include brands, offers, organizations, and organizational activities other than purchases. As a result, according to Zhang et al. (2017), engaged consumer behaviors go beyond simple transaction and consumption.

Engagement is a multidimensional construct with cognitive, emotional, and behavioral dimensions (Hollebeek, 2011a, b, 2013; Hollebeek et al., 2014; So et al., 2014, 2016), however the specific expression of these generic dimensions can vary depending on the idea of engagement used, and the relative importance of each one in relation to the context (So et al., 2014, Brodie et al., 2013; Hollebeek et al., 2014).

(Yoshida et al.,2014) defined fan engagement in sport industry as “a sport consumer’s extraoral behaviors in non-transactional exchanges to benefit his or her favorite sport team, the team’s management, and other fans”. Examples of customer engagement in sport industry could be Attending a game, watching it on TV, buying team merchandise, following their team on social media, and communicating with others about their team are all examples of sports fan involvement (Yoshida et al.,2014). (Hollebeek, 2011) described customer engagement as “the level at which the motivational, brand-related, and context-dependent state of mind of a customer is expressed, in terms of a degree of activation, identification, and absorption in brand interactions”.

Decision Making in Social Media

Recent studies have shown contributions offered by the literature on the relevance and the impact of sociality in the decision processes (Del Giudice, M, Caputo, F, & Evangelista, E., 2016; Yim, B, Baker, T, Byon, K& Zhang, J, 2020). Some researchers have examined the impact of user-generated content on personal, group and managerial decision-making (Power & Phillips-Wren, 2011). Other ttudies such (Luckshay,2019) suggests that the SMC impacts the decision-making by creating more word of mouth and retrieve some related information. Customers assume that opinions and believes on SMC can increase the pressure on decision making process. In addition, (Power & Phillips-Wren, 2011) research emphasizes that customers opinions and believes can impact businesses managers. This area of research is seeking to understand the influence that individuals make on managers in real-time communications in social media. According to (Radick, 2010) businesses, social media technologies provide unrivaled prospects.

Social media has been "important in boosting consumer awareness and enabling access to massive volumes of information, which impacts decision-making processes,"

The 12 Key Findings on Social Media's Impact on Business and Decision-Making by CEOs and Managers from Bulmer and DiMauro's survey of 356 professionals were as follows:
1) Professionals are more likely to be members of various social media platforms for commercial objectives. 2) LinkedIn, Facebook, and Twitter, the "Big Three" social media platforms, have risen to prominence as professional platforms. 3) Mobile is becoming a popular way for professionals to connect. 4) Social media is disrupting traditional decision-making procedures. 5) Professional networks are becoming a more important decision-making tool. 6) Information collected through internet networks has a high level of trust. 7) The usage of social media by companies, both internally and externally, is changing. 8) The value of peer feedback in decision-making is well acknowledged. 9) Professionals' usage of social media is driven by the desire to connect and collaborate. 10) "Final" decision makers are more likely to say they use a search engine to do their research. 11) Those with larger networks are more inclined to obtain information from their online network, read blogs, and query the Twitter channel as early as possible in the decision-making process. 12) When compared to older groups, younger responders are more likely to read a corporate blog and question the Twitter channel.

To measure decision-making performance efficiency and effectiveness. This study will use the scale that obtained from earlier research by (Gable et al., 2008; Huber, 1990; Tippins & Sohi, 2003). DMP includes 8 items scale which are "Organizational communication for effective decision making," "culture of long-term planning," "effective decision making," "speed in analyzing information," "time management for decision making," "reaching accurate and comprehensive information," "rapid and accurate identification of problems and opportunities," and "delegation of decision making". A 5-items Likert scale is going to be used from totally disagree (1) to totally agree (5).

THORETICAL BACKGROUND AND RESEARCH QUESITONS

The media richness theory and media synchronicity theory provide the theoretical foundations for this study.

According to media richness theory social views, message accuracy, and capacity to judge others, influence how media richness affects choice quality (Daft and Lengel, 1984). Richer media enhance social perceptions and the apparent capacity to detect dishonesty and knowledge in others. When participants have less task-relevant expertise, tools like electronic mail and electronic conferencing help to clarify communication. Kahai and Cooper (2003) discovered that the levels of participant competence and deception influence mediating constructs on decision quality. In general, when participants' task-relevant knowledge is high, richer media has a considerable beneficial influence on decision quality. Furthermore, by using richer media, the consequences of participant deception can be reduced. According to the media richness theory, an individual's perception of media richness is ascertained by the media's ability to provide immediate feedback, send multiple natural remedies via multiple communication channels, communicate in a variety of languages, and deliver personalized messages (Fernandez et al., 2013), all of which influence their media selection and use (Timmerman, 2002).

Media synchronization theory, according to (Windeler, 2018), predicts communication and task performance based on the capabilities of a communication medium. According to MST, media have a set of skills that are appropriate for specific communication processes. According to MST, when communication satisfies the condition of media fit, which arises when the medium's ability to promote synchronization matches the communication process and appropriation variables, communication and task performance improve (Lam, 2016)

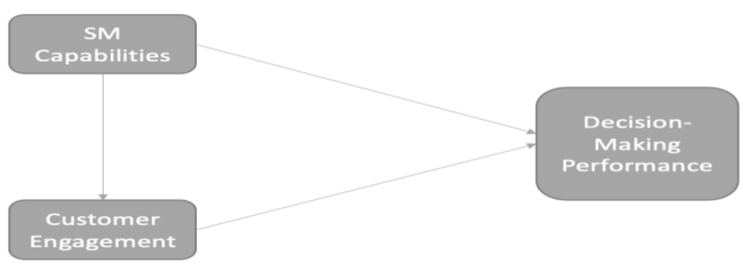
Synchronicity is defined as ‘‘a shared pattern of coordinated behavior as [teammates] work together’’ (Dennis et al., 2008). Five fundamental media capabilities determine synchronicity: transmission velocity, parallelism, symbol sets, rehears ability. Highly synchronous mediums have higher transmission velocity and more natural symbol sets but have less parallelism, rehears ability, and reprocess ability.

This study suggests media richness theory and media synchronicity theory for integrating several SMC and customer engagement dimensions, their cooperative effects on decision-making performance.
This paper answers the following research questions:

1. To what extent does social-media capabilities associated with decision-making performance within sport organizations?
2. How does customer engagement impact decision-making performance?

RESEARCH CONCEPTUAL FRAMEWORK

This study proposes a research model that interrelates social media capabilities, customer engagement, and decision-making performance variables as shown below.



RESEARCH HYPOTHESES

- This current study examines the relationship between decision-makers and fans the following hypotheses:
- **<u>H1</u>**: Social-Media capabilities will be positively associated with decision-making performance in sport organizations.
- **<u>H2</u>**: Customer Engagement will be positively associated with decision-making performance in sport organizations.
- **<u>H3</u>**: Social-Media capabilities will be associated with positively with customer engagement.

METHODOLOGY

Measures
Decision-making performance

To examine decision-making performance efficiency and effectiveness. This study will use the scale that obtained from earlier research by (Gable et al., 2008; Huber, 1990; Tippins & Sohi, 2003). DMP includes 8 items scale which are "Organizational communication for effective decision making," "culture of long-term planning," "effective decision making," "speed in analyzing information," "time management for decision making," "reaching accurate and comprehensive information," "rapid and accurate identification of problems and opportunities," and "delegation of decision making". A 5-items Likert scale is going to be used from totally disagree (1) to totally agree (5).

FUTURE STUDIES & LIMITATION

Examining the cultural difference of stakeholders and fans engagement on social media is considered as a key factor in these types of studies. In 2010, Kim, Y, Sohn, D, & Choi, S. study emphasis the cultural difference in motivations for using social network sites in Korea and America. Therefore, identifying how cultural differences impact social media use is a considerable factor to be investigated in future studies.

References

Abeza, G., O'Reilly, N., Séguin, B., & Nzindukiyimana, O. (2015). Social media scholarship in sport management research: A critical review. *Journal of Sport Management*, *29*(6), 601-618.

Al-Qaysi, N., Mohamad-Nordin, N., & Al-Emran, M. (2020). A systematic review of social media acceptance from the perspective of educational and information systems theories and models. *Journal of Educational Computing Research*, *57*(8), 2085-2109.

Al-Qaysi, N., & Al-Emran, M. (2017). Code-switching usage in social media: a case study from Oman. *International Journal of Information Technology and Language Studies*, *1*(1), 25-38.

Luckshay Arora, N. (2019). Analysis of the Impact of Social platforms on Decision-Making. *Journal of the Gujarat Research Society*, *21*(11), 393-397.

Barnard, S. R. (2016). 'Tweet or be sacked': Twitter and the new elements of journalistic practice. *Journalism*, *17*(2), 190-207.

Bagozzi, R. P., Dholakia, U. M., & Mookerjee, A. (2006). Individual and group bases of social influence in online environments. *Media Psychology*, *8*(2), 95-126.

Bruns, A., Highfield, T., & Lind, R. A. (2012). Blogs, Twitter, and breaking news: The produsage of citizen journalism. *Produsing theory in a digital world: The intersection of audiences and production in contemporary theory*, *80*(2012), 15-32.

Del Giudice, M., Caputo, F., & Evangelista, F. (2016). How are decision systems changing? The contribution of social media to the management of decisional liquefaction. *Journal of Decision systems*, *25*(3), 214-226.

Gruettner, A., Vitisvorakarn, M., Wambsganss, T., Rietsche, R., & Back, A. (2020, January). The New Window to Athletes' Soul–What Social Media Tells Us About Athletes' Performances. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

He, W., Tian, X., Chen, Y., & Chong, D. (2016). Actionable social media competitive analytics for understanding customer experiences. *Journal of Computer Information Systems*, *56*(2), 145-155.

Kim, Y., Sohn, D., & Choi, S. M. (2011). Cultural difference in motivations for using social network sites: A comparative study of American and Korean college students. *Computers in human behavior*, *27*(1), 365-372.

Kim, Y., Kim, Y., Wang, Y., & Lee, N. Y. (2016). Uses and gratifications, journalists' Twitter use, and relational satisfaction with the public. *Journal of Broadcasting & Electronic Media*, *60*(3), 503-526.

Power, D. J., & Phillips-Wren, G. (2011). Impact of social media and Web 2.0 on decision-making. *Journal of decision systems*, *20*(3), 249-261.

Toma, M. (2017). Missed shots at the free-throw line: Analyzing the determinants of choking under pressure. *Journal of Sports Economics*, *18*(6), 539-559.

Roth, P. L., Bobko, P., Van Iddekinge, C. H., & Thatcher, J. B. (2016). Social media in employee-selection-related decisions: A research agenda for uncharted territory. *Journal of management*, *42*(1), 269-298.

Roy, S. K. (2009). Internet uses and gratifications: A survey in the Indian context. *Computers in Human Behavior*, *25*(4), 878-886.

Singh, N. (2020). Sport Analytics: A Review. *The International Technology Management Review*, *9*(1), 64-69.

Sanderson, J. (2013). From loving the hero to despising the villain: Sports fans, Facebook, and social identity threats. *Mass Communication and Society*, *16*(4), 487-509.

Yim, B. H., & Byon, K. K. (2020). Critical factors in the sport consumption decision making process of millennial fans: a revised model of goal-directed behavior. *International Journal of Sports Marketing and Sponsorship*.

Yim, B. H., Byon, K. K., Baker, T. A., & Zhang, J. J. (2020). Identifying critical factors in sport consumption decision making of millennial sport fans: mixed-methods approach. *European Sport Management Quarterly*, 1-20.

Wang, Y., & Yu, C. (2017). Social interaction-based consumer decision-making model in social commerce: The role of word of mouth and observational learning. *International Journal of Information Management*, *37*(3), 179-189.

Zhi, Q., Lin, S., Talkad Sukumar, P., & Metoyer, R. (2019, May). GameViews: Understanding and Supporting Data-driven Sports Storytelling. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Bulmer, D., & DiMauro, V. (2011). 2nd Annual New Symbiosis of Professional Networks study. *Retriewed from http://dssresources. com/news/3255. php*.

Rubin, A. M. (2002). The uses-and-gratifications perspective of media effects. *Media effects: Advances in theory and research*, *2*, 525-548.

Radick S. 2010. The Getting Started with Government 2.0 Guide; [cited 2011 Jul 19]. Available from: http://stever- adick.com/2010/04/02/the-%E2%80%9Cgetting-started- with-government-2-0%E2%80%9D-guide/

Sashi, C. M. (2012). Customer engagement, buyer-seller relation- ships, and social media. *Management Decision*, *50*, 253–272. https://doi.org/10.1108/00251741211203551

Lim, J. S., Hwang, Y., Kim, S., & Biocca, F. A. (2015). How social media engagement leads to sports channel loyalty: Mediating roles of social presence and channel commitment. *Computers in Human Behavior*, *46*, 158–167.

Yoshida, M. and James, J.D. (2010), “Customer satisfaction with game and service experiences: antecedents and consequences”, Journal of Sport Management, Vol. 24, pp. 338-361.

Lam, C. (2016). Improving technical communication group projects: An experimental study of media synchronicity theory training on communication outcomes. *Journal of Business and Technical Communication*, *30*(1), 85-112.

Dennis, A. R., Fuller, R. M., & Valacich, J. S. (2008). Media, tasks, and commu- nication processes: A theory of media synchronicity. MIS Quarterly, 32, 575–600.

DECISION SCIENCES INSTITUTE

Sulaco decision framework of cross-functional collaboration

Tom Hall
University of South Florida
tchall@usf.edu

ABSTRACT

Sulaco Ventures, a B2B data company, saw the need for the creation of a cross-functional business unit to streamline decision making and combat product release delays. The assumption was that teams were making decisions in silos and not communicating with teams they had little to no direct impact on. A formalized decision framework was designed that created an entirely new position within the company to increase communication amongst the 30 or more cross-functional teams. This paper looks at the case's background causes and outcomes, explores the frameworks and relevant theories, and discusses the need for further research into the topic area.

KEYWORDS: Product Management, Cross-Functional Collaboration, Decision Making, Technology, Communication, Innovation

INTRODUCTION

In 2019 Sulaco Ventures, a leader in data management solutions, was missing its target product release dates 51% of the time and by an average of 52 days after the committed date. Being a non-public B2B company, Sulaco didn't face the same pressures a public and/or B2C company would to release its products on time, but leadership knew with the changing landscape of the tech industry that it had to be proactive to combat release delays. A cross-functional collaboration program was created called VPL (Ventures Program Leadership) that would drastically reduce the number and duration of delays hitting product releases. The company quickly saw improvements with achieving product release goals, and drastically reduced average delays, which can be seen in table 1 below. Even with these improvements, Sulaco needed to continuously evolve the program as the company's release train model shifted to a more agile, on-demand approach for both software and hardware releases. Additionally, supply chain issues with vendors had caused internal delays due to the lack of communication between different teams with second-sourced components being incorporated and qualified into Sulaco's products without fully complete documentation and support enablement.

Unlike other firms, who often mold their product releases to their program management process, Sulaco took the opposite approach and made the decision to mold their VPL program to the product releases, adapting to constant changes necessitated by the industry. By doing so, they were able to be more resilient to system shocks such as supply chain constraints, labor shortages, or release cadence changes, while still hitting their stretch goals. It also forced all teams to continuously reassess their internal processes. The key variable in this process is what Sulaco called the Business Program Manager (BPgM). The role's purpose was to sit between the cross-functional teams (CFT) and drive communication and decision making towards a common goal. The role itself was unique because it valued relationship building more than technical or business acumen by having someone in place who could manage the diverse personalities and departmental structures from a fully remote team that had members across the United States, Europe, China, and India.

Table 1: Annual Product Releases

	2019	2020	2021
Total Product Releases	47	61	53
Total on Time	23 (49%)	40 (66%)	46 (87%)
Total Late	24 (51%)	20 (33%)	4 (8%)
Average Delay	52 Days	28 Days	13 Days

Source: Sulaco Venture's Program Leadership Team.

VPL

Sulaco created the VPL program in 2019 with the goal of reducing the number and duration of product release delays. The program created the BPgM, who would facilitate cross-functional weekly meetings called Business Program Teams (BPTs) for each product. The BPTs would feature representatives from over 30 business units ranging from engineering to sales, finance, IT, legal, and more. The goal was to bring all the teams closer in alignment and to understand why teams did things a certain way and how those processes and decisions impacted other units.

Additionally, core teams would be created that featured the BPgMs, Engineering Program Managers (EPgM), Product Managers (PM), and technical team leads for each product, which would hold their own core team meetings weekly. The goal was to increase communication during the life-cycle of a product release to maintain visibility into the process and highlight any issues or concerns that teams might need assistance with. A minor issue for team X could have a major impact on team Y, but without a forum to discuss these issues they were at risk of going unnoticed until too late in the process.

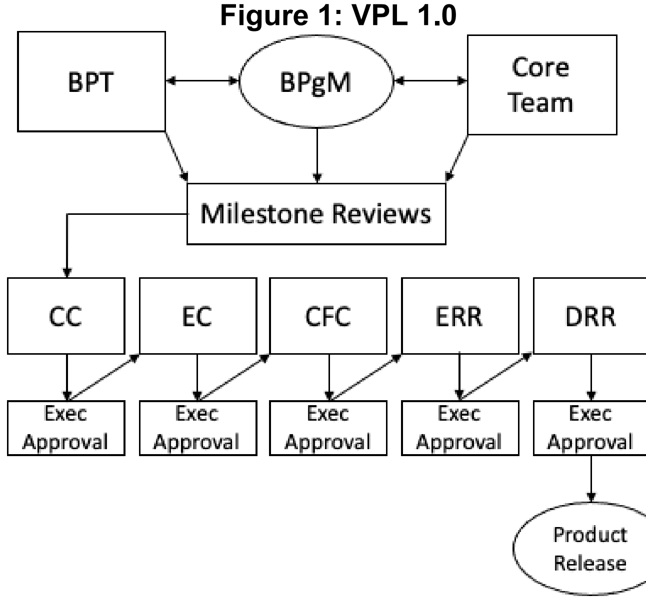
The result was VPL 1.0. Milestones were created that would provide specific content that would be compiled and reviewed by the BPTs and Core Teams before going in front of an executive approval team in the Executive Ventures Program Leadership (eVPL) meeting. They can be found below in Table 2 followed by the Figure 1 model of VPL 1.0. If approved, the product would move on to the next milestone before hitting general availability (GA), also known as the release date. Change control processes were in-place for any major shifts in dates or features, and post launch reviews were conducted with BPTs and Core Teams 90 days after the release.

Product releases typically occurred 1-2xs per year at this stage at Sulaco, which left substantial lead times, however the future plans were to move to an on-demand cadence which could see product releases occurring monthly or quarterly. Due to this changing release cadence, the process continued to become more streamlined with VPL 2.0 and 3.0, removing and combining stages so that there were less milestone reviews while empowering the various business units to make and own decisions such as process changes, approvals, and more.

Table 2: VPL 1.0

Milestone	Owner	Content
Concept Commit (CC)	Product Manager (PM)	CC is focused on the existing state of the product, competitive analysis, revenue, and desired features of the next release. Conducted as early as possible.
Engineering Commit (EC)	Engineering Program Manager (EPgM)	EC is focused on the engineering and technical aspects of the release based on content from the CC. Conducted within 2 weeks of the CC.
Cross-Functional Commit (CFC)	Business Program Manager (BPgM)	CFC is focused on alignment between the BPTs & Core Teams with the content approved at CC & EC. Conducted within 2 weeks of EC.
Execution Readiness Review (ERR)	Business Program Manager (BPgM)	ERR provides an update on the current status of releases to include the addition of quality testing as well as the plan of record (POR) and content for relevant members of the BPTs. Conducted at the 50% completion point for the release.
Delivery Readiness Review (DRR)	Business Program Manager (BPgM)	DRR is a pre-check that covers the final statuses of engineering and cross-functional team content. Final checkpoint for the release. Conducted 2-4 weeks before the final release.
General Availability	N/A	The release of the final product.

Figure 1: VPL 1.0

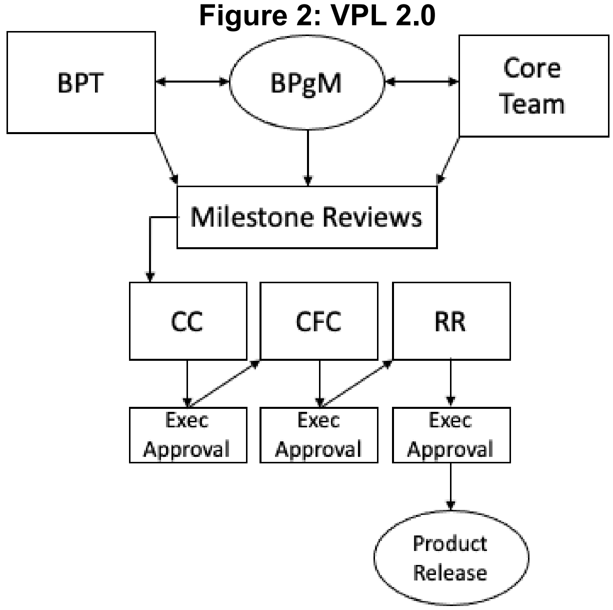


By the end of 2020 the VPL 1.0 program saw the desired results, with a 17% reduction in delayed releases and 24-day reduction in average delay times. The program was a success. However, despite the improvements Sulaco had still only hit 66% of its targeted release dates. Thus, an overhaul to the VPL process was conducted, with the result being VPL 2.0 in 2021.

Table 3: VPL 2.0

Milestone	Owner	Content
Concept Commit (CC)	Product Manager (PM)	Combines CC & EC from VPL 1.0. The goal is to highlight long-term & short-term strategy for the product to include competitive analyses, revenue pipelines, and resource investments. Conducted as early in release as possible.
Cross-Functional Commit (CFC)	Business Program Manager (BPgM)	Combines the CFC & ERR from VPL 1.0. The goal is to ensure engineering and business units are aligned on the committed plan, identifies all risks, and sets expectations for validations, demos, and more. Conducted 5-6 months prior to product release date.
Release Readiness (RR)	Business Program Manager (BPgM)	Maintains the role of the DRR from VPL 1.0 and acts as a final check with all cross-functional stakeholders to recommend go or no-go for release. Upon approval the product is cleared to release. Conducted one month prior to targeted product release.
General Availability	N/A	The release of the final product.

Figure 2: VPL 2.0

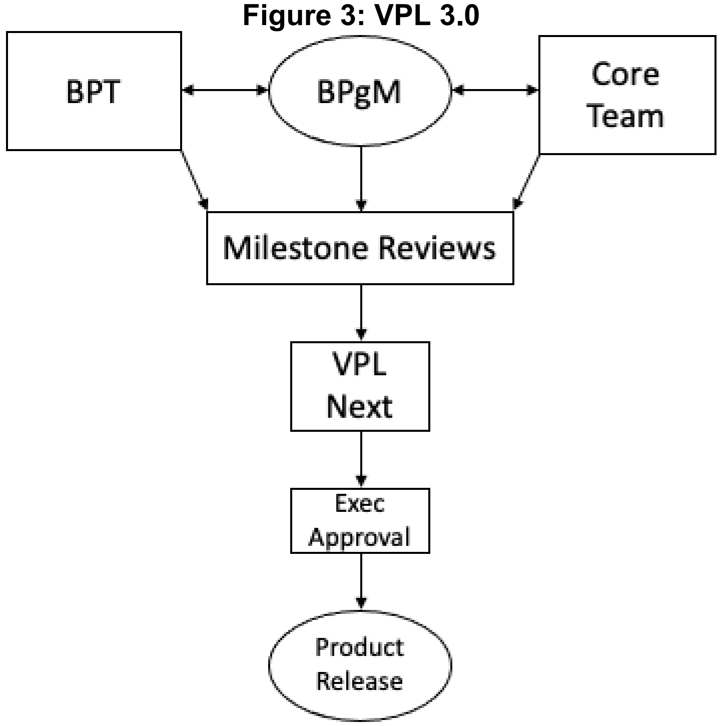


The new 2.0 iteration of the VPL process saw additional improvement, reducing the delayed releases by a further 21% and dropping the average delay by another 15 days. Despite these improvements, Sulaco continued to assess the efficiency of the VPL program. With the planned shift of the release cadence to become faster and release to occur more often, the process needed further adaptation to work effectively with future product releases. This is what led to the VPL 3.0 iteration in 2022, with the eventual rollout in 2023.

Table 4: VPL 3.0

Milestone	Owner	Content
VPL Next	Product Manager (PM) Engineering Program Manager (EPgM) Business Program Manager (BPgM)	Combines the CC, CFC, and RR from the VPL 2.0 process. The goal is to reduce process overhead while maintaining a view of risks, focus on short-term goals and long-term strategy, and prioritize the feature backlogs to align with the strategy. Primary decision making is shifted from the executive level to the cross-functional team level. Conducted 45-60 days prior to the product release.
General Availability	N/A	The release of the final product.

Figure 3: VPL 3.0



Literature Review & Theories

The idea behind the benefits of cross-functional communication to better drive work processes is well-known. Hackman (1986) is one of the leaders in this field with the Group Effectiveness Model, which highlights that a group needs to be a team, needs compelling direction, must have an enabling structure and supportive organizational context, as well as expert coaching. Additional people have contributed to this model, such as Katzenback & Smith (1993), LaFasto & Larson (2001), and Lencioni (2002). Similar models have also been created elsewhere such as the Contingency Theory Model (Fielder, 1964) and the Cooperative model of Knowledge Sharing (Ghobadi & D'Ambra, 2012; Loebecke et al, 1999; Van de Ven, 1976).

Businesses and governments have also taken to creating their own CFC models, such as Korn Ferry's T7 model (Meuse, 2009). Others have created cultures instead of processes, such as Gore Associates, the makers of Gore-Tex, who have eliminated major silos within their company by creating a culture where each facility has no more than 150 employees, and all titles are removed with the goal of encouraging conversations amongst workers from different departments (Gladwell, 2002). Publicly, General Stanley McChrystal outlined in his book *Team of Teams* how Joint Special Operations Command (JSOC) broke down silos and increased cross-functional communication to decimate Al Qaeda in 2005/2006 after struggling to do so in the earlier phase of the war, empowering special operations forces to communicate and work with other units and branches to accomplish a shared goal (McChrystal, 2015). In both examples, the programs and processes were adapted to fit their use case or products.

When looking at CFC applications throughout the technology sector, the common themes are silos being broken down and communication being improved at a team, project, or product level (Cascade, 2022). Dell created a cross-functional team that focuses strictly on its Apex portfolio (Haranas, 2022), RedHat uses cross-functional technical teams to facilitate configuration management (Kumar & McMillen, 2020), and Seagate utilized cross-functional teams for the Optical Inspection and Centralized Analysis (OPICA) project that won it the National Association of Manufacturers 2022 Manufacturing Leadership Award (Business Wire, 2022). There no doubt exists programs like Sulaco's in other firms, but with firms having no motivation or reason to share their internal processes publicly, it is difficult to gain that understanding from literature and more research is warranted.

The theories and practices articulate that companies need to go beyond communication. Having reliable processes is just as vital to a project, program, or product as communication. By having processes to follow for given situations and benchmarks, Sulaco gave all cross-functional team members an understanding of their left and right limits. For example, if the release date has to be moved from the agreed upon date, the teams know they need to initiate a Change Request to the executive team for approval. Without the established process, the executive team might not be aware of the release date change or members of the cross-functional team might proceed forward unaware of the change, thus causing backend issues. The decision to push a release could also be disagreed with or overruled by the executive team.

VPL 1.0 was able to create and implement these processes with an agile perspective that saw Sulaco continuously reassessing and reevaluating the efficiencies and impact of the program. By taking in feedback and solving pain points, the VPL process became malleable where other organization's processes might remain rigid. Figure 4 shows what we will dub the Sulaco Framework of cross-functional collaboration that can be company agnostic. This framework includes a post-release review, which Sulaco currently does, that sources feedback from all the teams involved in the release via survey to make improvements to processes.

Table 5: Sulaco Decision Framework

Step	Owner	Explanation
Milestone Review	Business and Technical Teams	The milestones need to be established to gain an understanding of the needed cadence and approval stages of the product's lifecycle. There is number of milestones is dependent on the product and process.
Executive Sponsor Approval	Business Program Manager (BPgM)	This stage might be a manager, a director, or a vice president; however the milestone needs to be socialized with higher leadership to prevent conflicting processes or misalignment of strategy.
Product Release	Business and Technical Teams	The release of the product occurs once the milestones are approved, and the product has reached it all committed requirements.
Post-Release Review	Business Program Manager (BPGM	The BPgM needs to consistently seek feedback from all the cross-functional teams to improve the decision making and processes on future product releases. This feedback is how Sulaco evolved from VPL 1.0 to 3.0

Figure 4: Sulaco Decision Framework

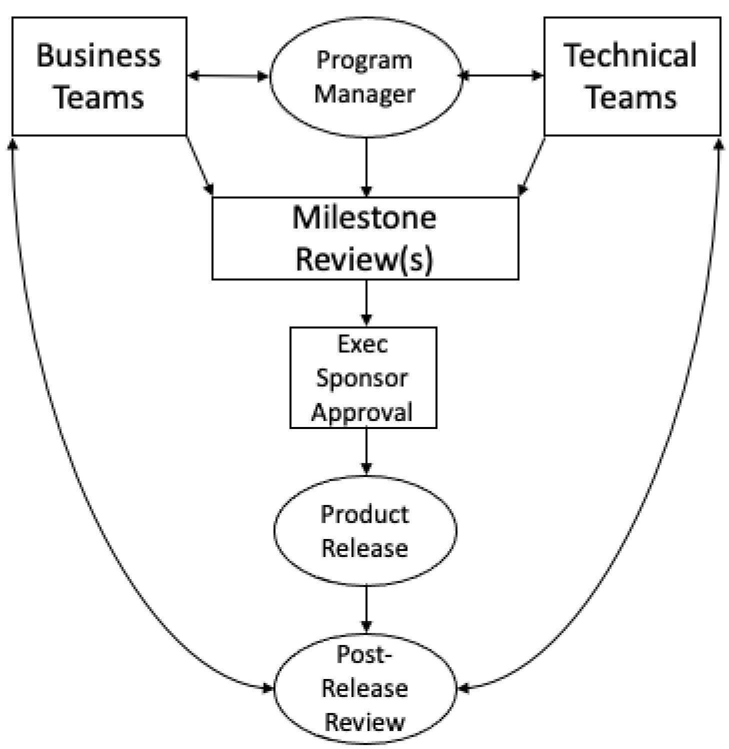


Table 6: Relevant Theories

The Group Effectiveness Model (GEM)	
● Multiple models with various names have been created, with the core principles involving trust, team leadership, team commitment, organizational culture and processes, and conflict management. ● Hackman is a leader in this field and proposes five conditions that lead to team effectiveness. ○ Real Team ○ Compelling Direction ○ Enabling Structure ○ Supportive Organizational Context ○ Expert Coaching ● Other researchers such as Rubin, Lovnick, and Fry, with their GRPI Model of Team Effectiveness, propose a top-down model in the form of a pyramid with four key levels: ○ *Goal*: Clearly defined and agreed upon by stakeholders. ○ *Roles*: Team leader picked and roles understood. ○ *Processes*: Workflow and team processes are in place. ○ *Interpersonal Relationships*: Trust, collaboration, and conflict management.	Hackman (1986) Hackman (2002) Katzenback & Smith (1993) LaFasto & Larson (2001) Lencioni (2002) Rubin, Plovnick, & Fry (1977)
Contingency Theory / Model	
● Contingency Theory is in many ways similar to Group Effectiveness but theorizes that there is no set “best way” to structure or organize a business. Instead, leaders play the key role of organizational effectiveness based on internal & external factors. ● Fiedler built upon this theory with the Contingency Model of Leadership Effectiveness. ○ Situational Favorableness ■ Affective Leader - Group Relations ● Trust amongst the team. ■ Task Structure ● Clear & detailed or ambiguous. ■ Position Power ● Authority of the group leader. ● Shepherd & Hougland look at contingency theory from the perspectives of *complex man* vs *complex organization.* ○ *Complex Man* ■ People’s social behaviors are developed through socialization in non-work environments. ○ *Complex Organization* ■ Business units and processes are complex, all interacting with each other in different ways.	Donaldson (2001) Fiedler (1964) Hersey, Blanchard, & Natemeyer (1979) Luthens & Stewart (1977) Montiz Jr (2010) Shephard & Hougland (1978)

Table 6 continued

Cooperative Model of Knowledge Sharing	

● Information and knowledge can be looked at as a commodity in an organization. Someone who possesses either can hold power within an organization and may not want to share it with a cross-functional team due to loyalty to their functional team.
    ○ Cooperation and competition impact knowledge sharing.
● *Coopetition*
    ○ Existence of cooperation and competition simultaneously.
    ○ Less relevant for internal cross-functional teams, applies more to cooperation amongst competing companies.
● Loebecke's Three Dimensions of Knowledge Transfer
    ○ *Synergy*
        ■ When sharing knowledge with the team creates more value than if the member holds the knowledge as an individual.
    ○ *Leveragability*
        ■ The amount at which a team receiving the knowledge can leverage it to create the additional value.
    ○ *Negative Reverse Impact*
        ■ The point where knowledge sharing could potentially cause negative effects when shared compared to if held by an individual.
● *Organizational Coordination*
    ○ Van de Ven says that organizational coordination "means integrating or linking together different parts of an organization to accomplish a collective set of tasks."
	Ghobadi & D'Ambra (2012)
	
	Loebecke et al (1999)
	
	Tsai (2002)
	
	Van de Ven (1976)

FUTURE RESEARCH

While the theories mentioned have been widely studied, the practical role of processes within a cross-functional program have not. There is a gap in the literature where the theory and practice should meet. Sulaco started this program from scratch, but had there been an existing framework and research, the positive results may have been gained far earlier in the process. The challenge is creating a study and/or framework that is useful and adaptable by businesses or organizations in multiple industries. The Gore Associates model wouldn't work on Sulaco and the Sulcao VPL framework might not be applicable to Gore Associates. With different cultural, organizational, and product factors at play, these gaps are what future research could help solve.

CONCLUSIONS

By checking the blocks of having a strong program leader in place, processes built out, and a supportive organizational culture, Sulaco met all the core criteria according to the Group Effectiveness Model and Contingency Theory. Sulaco also left the program relatively flat, with a small number of key decision makers able to quickly test and make changes to the program or individual processes to evolve with product release cadences whether they be annually, quarterly, or monthly. VPL 1.0 saw a 17% improvement to product release delays, but by continuously assessing the program and processes they were able to further reduce the delays another 21% in VPL 2.0, all while maintaining their product release cadences.

However, the program as it exists in VPL 3.0 is not without limitations. Each iteration brings with it new challenges, such as the concern of socializing committed product release features at a short notice of 45-60 days prior to the GA date, whereas in VPL 1.0 and 2.0 they were socialized 4-5 months in advance, giving the cross-functional teams and executive leadership substantial ramp up time. These types of problems will arise with each new iteration. This increases the importance of having a supportive organizational culture and strong leader, which Sulaco does, as the BPgMs are empowered to adjust to fill gaps left by advancing to new iteration of the process. This valuable feedback and data from the BPgMs are documented and then utilized to help evolve the VPL process, which is currently driving for a 4.0 release by 2024.

Author’s Note

At the request of the company’s legal team, the firm’s name was changed to Sulaco Ventures in the event that the paper is published. All stats, figures, and storylines are accurately represented in this case study with permission from the company received. For further information, please contact the author directly.

REFERENCES

Business Wire (2022, March 31). *Seagate Wins Manufacturing Leadership Award from a Jury of Its Peers*. https://www.businesswire.com/news/home/20220331005430/en/

Cascade Team (2022, April 14). Top 4 Cross Functional Collaboration Examples. https://www.cascade.app/blog/cross-functional-teams-drive-innovation

De Meuse, K. P. (2009). Driving Team Effectiveness. Korn/Ferry Institute. https://jocon.com/fileadmin/download/pdf/Teamswhitepaper.pdf

Donaldson, L. (2001). The contingency theory of organizations. Sage.

Fiedler, F. E. (1964). A contingency model of leadership effectiveness. In Advances in experimental social psychology(Vol. 1, pp. 149-190). Academic Press.

Gerstberger, W. (2021). The Balance of Cross-Functional Collaboration and Stretch and its Interaction with Firms' Innovation Performance. Journal of Innovation Management, 9(3).

Ghobadi, S., & D'Ambra, J. (2012). Knowledge sharing in cross functional teams: a coopetitive model. Journal of Knowledge Management.

Ghobadi, S., & D'Ambra, J. (2013). Modeling high-quality knowledge sharing in cross-functional software development teams. Information Processing & Management, 49(1), 138-157.

Gladwell, M. (2002). The tipping point. Back Bay Books.

Hackman, J. R. (1986). The Design and Management of Work Teams. YALE UNIV NEW HAVEN CONN.

Hackman, J. R. (2002). Leading teams: Setting the stage for great performances. Boston: Harvard Business School Press.

Haranas, M. (2022, March 7). Dell Creates New Partner-Focused Apex 'Cross-Functional Team'. CRN. Dell Creates New Partner-Focused Apex 'Cross-Functional Team'

Hersey, P., Blanchard, K. H., & Natemeyer, W. E. (1979). Situational leadership, perception, and the impact of power. Group & organization studies, 4(4), 418-428.

Jeske, D., & Calvard, T. S. (2020). A review of the literature on cross-functional integration (2010–2020): trends and recommendations. International Journal of Organizational Analysis.

Katzenbach, J. R., & Smith, D. K. (1993). The wisdom of teams: Creating the high-performance organization. Boston: Harvard Business School Press.

LaFasto, F., & Larson, C. (2001). When teams work best: 6000 team members and leaders tell what it takes to succeed. Thousand Oaks, CA: Sage.

Lencioni, P. (2002). The five dysfunctions of a team: A leadership fable. San Francisco: Jossey-Bass.

Loebecke, C., Van Fenema, P. C., & Powell, P. (1999). Co-opetition and knowledge transfer. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 30(2), 14-25.

Luthans, F., & Stewart, T. I. (1977). A general contingency theory of management. Academy of management Review, 2(2), 181-195.

McChrystal, G. S. A., Silverman, D., Collins, T., & Fussell, C. (2015). Team of teams. Portfolio Penguin.

McMillen, D., & Kumar, V. (2020, September 9). Red Hat on Red Hat: How collaboration can transform configuration management in IT. Red Hat on Red Hat: How collaboration can transform configuration management in IT

Moniz Jr, R. (2010). Practical and effective management of libraries: integrating case studies, general management theory and self-understanding. Elsevier.

Rubin, I. M., Plovnick, M. S., & Fry, R. E. (1977). Task oriented team development. New York: McGraw-Hill.

Shepard, J. M., & Hougland Jr, J. G. (1978). Contingency theory:“complex man” or “complex organization”?. Academy of Management Review, 3(3), 413-427.

Tsai, W. (2002). Social structure of “coopetition” within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. Organization science, 13(2), 179-190.

Van de Ven, A. H., Delbecq, A. L., & Koenig Jr, R. (1976). Determinants of coordination modes within organizations. American sociological review, 322-338.

DECISION SCIENCES INSTITUTE

Teaching Analytics without Coding

Yao Zhao
Rutgers University
Email: yaozhao@rutgers.edu

ABSTRACT

Many students like analytics, but few like coding. We introduce a web-based business database, SCDATA.AI, integrated with data models and visualization tools, to teach a variety of decision science and analytics topics, such as job intelligence, inventory analytics, sourcing analytics and competitive intelligence, without coding. Within a few clicks on a graphic user interface, students can access and play with a vast dataset of 30,000+ companies from the world's top 20+ stock exchanges (about 70% of Standard & Poor's active companies), to learn how to generate, interpret and present business insights.

KEYWORDS: Teaching Analytics, Data Visualization, Business Database, Job Intelligence, Inventory Analytics, Sourcing Analytics, Competitive Intelligence, Coding.

DECISION SCIENCES INSTITUTE

Teaching Blockchain in USA Universities

Teuta Cata
Northern Kentucky University
Email: catat@nku.edu

Susan Brudvig
Northern Kentucky University
Email: brudvigs1@nku.edu

Xiaoni Zhang
University of Alabama at Birmingham
Email: xzhang2@uab.edu

ABSTRACT

Given the growing importance of blockchain technologies, we examine the response by US universities. Specifically, we examine the degree to which US universities offer courses covering blockchain content and offer extracurricular activities about blockchain technologies. Our preliminary results find that 25% of US universities provide content coverage of blockchain in coursework. However, the preponderance (57%) of US universities offered extracurricular activities, suggesting that most universities expose students to emerging technologies outside the classroom.

KEYWORDS: Blockchain, Crypto, Curriculum, Micro-Credentials, Extracurricular Activities

INTRODUCTION

The need for blockchain professionals has increased dramatically. The market for blockchain technologies is increasing at a compound annual growth rate of 36%, and the demand for blockchain engineers has increased by more than 500% in 2022 alone (Patan et al., 2023). In a survey of senior executives, over half indicated that blockchain skills were a critical priority for their organizations (Deloitte, 2019). However, the education system has not kept pace with providing qualified professionals, leading to a shortage in the job market.

The need for more highly qualified professionals with blockchain expertise can be addressed with different approaches. For instance, traditional universities can provide new blockchain courses or blockchain content in existing courses. Another approach would be to offer extra-curricular training for graduates and undergraduate students where they can learn blockchain concepts and skills currently practiced in the industry. Blockchain micro-credentials also can be offered online, targeted to students with limited access to on-campus instruction, whether domestically or abroad (Jirgensons & Kapenieks, 2018). In addition to the instruction offered by traditional universities, well-known education platforms such as Coursera, EdX, and FutureLearn offer courses and certificates in blockchain topics (Holotescu et al., 2018). Finally, students can participate in organizations independent of institutions. For example, over 3000 students from 60 countries participate in the Blockchain Education Network, a way for students to get involved in blockchain education. While it is apparent that there are many approaches to addressing the need for qualified professionals, what needs to be clarified is the degree to which blockchain extracurricular activities and blockchain courses are offered by traditional universities.

In this study, we systematically study blockchain education in US universities to determine the breadth and depth of blockchain content. Our study provides insight into the response of the American higher education system to include emerging technologies in the curriculum. We view curricula broadly, and a unique contribution of our study is that we also include blockchain extracurricular activities in our review of blockchain in US universities.

Given the importance of blockchain technology in different industries and the various institutions teaching different aspects of such technology, the purpose of this study is to provide an overview of the status of blockchain curricula taught in universities in the United States. This research focuses on the following research questions:

RQ1: To what extent are blockchain extracurricular activities offered at US universities?
RQ2: To what extent do US universities offer blockchain courses?
RQ3: What types of programs and disciplines offer blockchain courses in US universities?

LITERATURE REVIEW

Blockchain technology, as a distributed ledger, was introduced by Satoshi Nakamoto in 2018. Blockchain has the potential to revolutionize existing systems by providing transparent, secure, and decentralized solutions, which are becoming increasingly important across various industries, including finance, supply chain management, healthcare, education, and more. There is a growing demand for blockchain professionals, including blockchain developers, consultants, analysts, and project managers, pushing universities to prepare students for high-paying job opportunities in the blockchain industry.

A recent survey conducted on blockchain challenges organizations face shows that the skills needed to develop and use blockchain applications are still in short supply (Brown, 2021). It is difficult to close the gap between blockchain jobs and blockchain skills, which has surged in the last couple of years. Traditional universities are behind in providing students with the skills needed for the digital economy. However, other non-academic organizations such as Udemy, RMIT University, 101 Blockchain offer various blockchain classes. Such classes vary from beginner to advanced, courses or certificates, hands-on or conceptual. Course fees vary widely, starting at $17 up to $23,000 per course (Patan et al., 2023). Patan and his colleagues (2023) suggest that Higher Education Institutions are lagging on blockchain course offerings and technical skills for such technology, resulting in an education/ job skills gap that will be very difficult to close soon.

Research in academia is focused on what would be the best approach to teaching blockchain classes. A systematic literature review conducted by Strang et al. (2020) used two dimensions to classify the research papers published on the topic of blockchain curriculum: Stakeholder Perspective (Managerial vs Technology Approach) and Teaching Ideology (Theoretical vs Applied). Authors indicate that blockchain technology, as a very modern cybersecurity technology should be taught by integrating risk management strategies, so students can understand the implications of such technology in the organization settings.

Teaching blockchain to students may have a slower approach than teaching IoT, which was an interesting topic to students. The best approach is to start teaching it at higher level courses or master's degrees, creating special topics, or slowly introducing blockchain concepts vis guest speakers, events, or campus clubs (Kursh & Gold, 2016). Blockchain as a novel topic can be introduced in the existing courses by having discussions about it, and having some hands-on exercises, which will make courses fun and interesting for students. The most well-suited disciplines to introduce blockchain are taught in Computer Science and Information Technology. Students can use Cloud applications to create, host, and run blockchain applications while learning about this technology. Cloud platforms offered by Google Compute Engine, Amazon AWS, or Microsoft Azure can help students to have access to such applications without using local and expensive hardware (Purdon & Erturk, 2017).

Blockchain technology is interdisciplinary by nature and combines elements of computer science, cryptography, economics, and law. Teaching blockchain encourages interdisciplinary collaboration and exposes students to a wide range of knowledge and perspectives. The most reputable United States universities, such as Harvard, MIT, and Duke, are teaching blockchain courses offered by a variety of departments, including Computer Science, Engineering, Business, Finance, and Law (Patan et al., 2023).

Blockchain classes in computer science cover various aspects of blockchain technology, including its fundamental concepts, architecture, protocols, consensus mechanisms, smart contracts, security, and applications. These classes typically focus on blockchain's theoretical and practical aspects, equipping students with the knowledge and skills necessary to understand, develop, and evaluate blockchain-based systems. Blockchain concepts can easily be integrated into Security classes since digital signatures and crypto hashing are security measures taught to computer science students (Shebaro, 2022). Other researchers have discussed the utilization of project-based learning in computer science classes where students programmed smart contracts using Solidity, expanding their knowledge of how blockchain can be used in different industries. Project-based learning helps students to learn blockchain concepts via hands-on projects and enables students to learn from peer's comments (Mentzer, et al., 2020). Shi et al. (2022) provide an active learning platform for computing students where they learn about blockchain via labs (such as proof of work concepts) and enhance their understanding of such new technology.

Blockchain teaching approach for business students involves providing them with a comprehensive understanding of blockchain technology and its applications in various business sectors. A survey of accounting professionals and professors suggests that blockchain technology should be taught to all business majors in Accounting Information Systems courses (Felski & Empey, 2020). Blockchain concepts taught to graduate accounting students via coding in R assignments increased the understanding of what blockchain is and how the internal structure of the blockchain is created. Such hands-on learning exercises increase the depth of understanding, which can't be delivered by just some conceptual lectures in class (Kaden et al., 2021). Interactive activities, such as hashing of proof of work and blockchain implementation in a supply chain case, increased the knowledge of the accounting students in blockchain technology without being a programmer (Stratapoulos, 2020).

A good discussion in the blockchain curriculum has been about how to teach blockchain and cryptos to business students. Dettling (2018) suggests that business students need to know the general concepts of blockchain, such as the new business models, smart contracts, the process redesign that is happening with the new technology; information on cryptocurrencies, such as different types of cryptos, the exchange platforms, and legal issues; building blocks of the technology such as cryptographic hashing, consensus algorithms, and distributed ledgers; and the applications of such technology in business settings such as in supply chain, eHealth, insurance, and eGovernment. Teaching materials for business students are not up-to-date or developed for a specific topic, making teaching the blockchain to such students a real challenge to faculty assigned to teach such courses.

Universities can include blockchain in their curricula by offering extracurricular activities, such as student clubs, student events, lecture series, and entrepreneurial ventures. A traditional approach is to create new courses. Universities are introducing blockchain as an overview in their curriculum or offering specialized blockchain courses; they include business use cases, research articles, guest speakers, and textbooks focused on blockchain and cryptos. Universities also can incorporate blockchain into existing courses. The broad variety of teaching materials and perspectives in this new technology is allowing instructors to supplement the teaching materials provided by the textbooks and give students broader information about blockchain and cryptocurrencies (Kursh & Gold, 2016).

METHODS

To address our research questions, data about blockchain courses and extracurricular activities were collected for US universities. The initial step in data collection involved creating a list of US universities. We used the publicly available website graduateshotline.com (https://university.graduateshotline.com/ubystate.html) to extract a list of US institutions of higher education by US state. We identified 880 US universities, after excluding colleges.

Data were collected from September to December 2022 by two graduate student coders. For each US university, the coders performed a keyword search of “Blockchain” on the university’s website. The following information was recorded for each US university:

> *Blockchain Extracurricular Activities* – Whether a university promoted any blockchain extracurricular activities, such as guest speakers and student events, as listed on the website, requesting students to attend or participate.
>
> *Blockchain Courses* – Course numbers, titles, and descriptions of any course offering blockchain or cryptocurrency in the course description; The college and department in which the course resides; The degree level of the course (graduate, undergraduate); The name of the degree, credential, or certificate for which the course applies, if applicable.

In addition, the US universities were categorized into regions as follows:

> *Northeast* – CT, DE, ME, MD, MA, NH, NJ, NY, PA, RI, VT
> *South* – AZ, AL, AR, FL, GA, KY, LA, MS, MO, NC, NM, OK, SC, TN, TX, VA
> *Midwest* – IL, IN, IA, KS, MI, MN, NE, ND, OH, SD, WI
> *West* – AK, CA, CO, HI, ID, MT, NV, OR, UT, WA, WY

In the final step, one of the authors reviewed the data for accuracy and consistency before the data was analyzed.

RESULTS

For this paper, our goal is to provide a preliminary analysis of the data collected, thereby addressing RQ1 to RQ3. In subsequent analyses, we intend to address our research questions more fully. For instance, our analytic approaches might include content analysis or text mining to identify the content of blockchain courses. We also intend to identify the disciplines in which blockchain courses are offered.

To address RQ1 and RQ2, the total number of universities was tallied, along with the number of universities offering block extracurricular activities and blockchain courses. This data was broken out by region of the United States. (See Table 1)

Table 1: US Universities with Blockchain Extracurricular Activities and Blockchain Courses

	US Universities	Blockchain Activities		Blockchain Courses	
	Total	Count	% Total	Count	% Total
Midwest	213	123	58%	37	17%
Northeast	148	90	61%	53	36%
South	362	178	49%	84	23%
West	157	107	68%	45	29%
Sum	880	498	57%	219	25%

Of 880 US universities, 498 (57%) offered blockchain extracurricular activities, indicating that 43% did not offer blockchain extracurricular activities. The highest proportion of universities offering blockchain extracurricular activities was in Western states (68%) and the lowest in Southern states (59%). A Chi-Square test examined the relationship between region and blockchain extracurricular activities. There was a statistically significant relationship between the region of a US university and whether blockchain extracurricular activities are offered, $\chi^2(3) = 17.84$, $p < 0.001$.

Of 880 US universities, 219 (25%) offered one or more blockchain courses. The highest proportion of universities offering a blockchain course was in Northeastern states (36%), and the lowest proportion was in Midwestern states (17%). A Chi-Square test examined the relationship between region and blockchain courses. There was a statistically significant relationship between the region of a US university and whether a blockchain course was offered, $\chi^2(3) = 17.63$, $p < 0.001$.

A correlation was prepared on two binary indicators to assess whether there is a correspondence between US universities offering blockchain extracurricular activities and a blockchain course. There was a statistically significant relationship between blockchain extracurricular activities and offering a blockchain course ($r = 0.446$, $p < 0.0001$). The correlation is moderate due to universities offering blockchain extracurricular activities but no blockchain courses (33%, $n = 287$). However, the converse is not true: Only 1% (n = 8) of US universities offered a blockchain course and no extracurricular activities.

A frequency table was prepared to examine the distribution of blockchain courses in US universities. Of 880 US universities, 660 (75%) did not offer a blockchain course. Of the 219 US universities offering blockchain courses, most universities (168 or 76%) offer just one course. (See Table 2)

Table 2: Blockchain Courses Offered in US Universities		
Number of Blockchain Courses Offered	Frequency	Distribution
0	660	75%
1	168	19%
2	29	3%
3	10	1%
4	8	1%
5 or more	5	1%
Total Universities	880	100%

To address RQ3, US universities' total number of blockchain courses (n = 880) was tallied, broken out by the number of undergraduate courses and graduate courses. Consistent with Table 1, this data was broken out by region of the United States. (See Table 3)

Table 3: Blockchain Courses in US Universities					
	Blockchain Courses	Undergraduate Courses		Graduate Courses	
	Total	Count	% Total	Count	% Total
Midwest	47	26	55%	21	45%
Northeast	65	41	63%	24	37%
South	134	74	55%	60	45%
West	76	51	67%	25	33%
Sum	322	192	60%	130	40%

Of the 322 blockchain courses US universities offer, 60% are at the undergraduate level, and 40% are at the graduate level. There are proportionally more undergraduate blockchain courses being offered in the West (67%) and Northeast (63%) and proportionally more graduate courses being offered in the Midwest (45%) and South (45%). However, the regional differences are not statistically significant, $\chi^2(2) = 3.53$, $p > 0.05$.

DISCUSSION AND CONCLUSIONS

The purpose of this study was to create an understanding of how blockchain technologies are being integrated into the curriculum in US universities. The motivation for the study was due to the growing need for blockchain professionals, coupled with an apparent lack of research describing how blockchain technology is being integrated into higher education. This study addresses this gap, and our conference paper provides a preliminary examination of the data.

Our results provide preliminary evidence of the degree to which blockchain technology is covered in at least one course. We find that 25% of US universities provide content coverage of blockchain in coursework, but the preponderance of institutions only offers a single course covering blockchain content. This suggests that only some institutions fully integrate blockchain into degree programs beyond the coverage of just one course. In future research, we intend to examine how blockchain integrates into different domains, such as information systems education (Leidig & Salmela, 2022).

We also identified the degree to which US universities offer blockchain extracurricular activities. Because more than half (57%) of US universities offered these activities, it suggests that most universities are exposing students to emerging technologies outside of the classroom. However, not all US universities provide this content to students. Specifically, our preliminary research suggests that 43% of US students (n = 374) provide neither blockchain courses nor blockchain activities. Future research could more fully understand these institutions' characteristics, comparing those who offer blockchain courses and/or extracurricular activities.

Finally, our results provide evidence of regional differences in the degree to which blockchain courses and extracurricular activities are offered in US universities. These results may mirror other studies identifying regional differences in emerging technologies (e.g., Zhao & Zhao, 2016). We intend to explore and understand the regional differences in future research.

REFERENCES

References are available upon request.

DECISION SCIENCES INSTITUTE

The Critical Success Factors and Barriers of Lean in Small and Medium-Sized Enterprises: A Systematic Literature Review

Fatma Pakdil
Eastern Connecticut State University
Email: pakdilf@easternct.edu

Kübra Gürtaş Türktekin
Middle East Technical University
Email: kgurtas@gmail.com

Arsev Umur Aydınoğlu
Middle East Technical University
Email: arsevu@gmail.com

ABSTRACT

Although Critical Success Factors (CSFs) are considered pivotal in a successful lean implementation, the literature shows limited studies that analyze the CSFs in depth. In light of this gap, this study explores the CSFs and barriers of lean production in Small and Medium Enterprises (SMEs). This study reveals that management commitment, leadership, employee commitment and engagement, and knowledge of lean tools are the most frequently used CSFs in the previous literature. This study also shows that the first four barriers are resistance to change, lack of training and knowledge, lack of leadership commitment and support, and lack of resources.

KEYWORDS: Lean manufacturing; critical success factors; Small and Medium Sized Enterprises

INTRODUCTION

Lean production (LP) has shown a growing interest in the manufacturing industry, as well as the service industry, over the last decades. Although the lean philosophy has allowed many firms to improve overall performance (e.g., Liker and Rother, 2011; Shah and Ward, 2003, 2007), research shows that a limited number of firms successfully implemented lean (Bhasin and Burcher, 2006; Pakdil and Leonard, 2014, 2017). Sohal and Egglestone (1994) stated that only approximately 10% of firms worldwide successfully implemented lean practices, while Repenning and Sterman (2001) found that 98 percent of such implementations do not create significant success.

Lean is a philosophy rather than a simple set of applications (Bhasin and Burcher, 2006). It is also known as the *just-in-time* (JIT) system because of its roots in the Toyota production system (TPS). The ability to reduce costs and increase flexibility also provides competitive advantages through superior firm performance in lean (e.g., Im and Lee, 1989; White et al., 1999; Shah and Ward 2003; Li et al., 2005; Narasimhan et al., 2006; Liker, 2004). LP focuses on identifying efficient and effective production systems that consume fewer resources, create higher quality, and reduce cost as outcomes (Pakdil and Leonard, 2014: 4588). Particularly, small and medium-sized enterprises (SMEs) have shown a pivotal interest in lean philosophy in the manufacturing industry worldwide. Given that more than 95% of registered firms worldwide are SMEs and SMEs generate more than 35% of the Gross Domestic Product (GDP) in emerging economies, SMEs are considered one of the driving forces of the national economies (Alibhai et al., 2017). Considering the unique features of the SMEs, literature shows that lean implementation in SMEs relies on some critical success factors (CSFs) and gets affected by the barriers and challenges. Although CSFs are considered pivotal in a successful lean implementation, the literature shows that there are limited studies analyzing CSFs in depth. Therefore, in light of this gap, the critical roles of SMEs, and growing interest in LP among SMEs, this study focuses on exploring the CSFs and barriers of LP in SMEs. Using the systematic literature review (SLR) process and bibliometric analysis, this study aims to reveal these CSFs and barriers to help SMEs identify what factors may improve the outcomes of lean implementation in SME context. To the best of our knowledge, the data synthesized and the knowledge generated through a SLR method in this study have not previously been presented in the relevant literature in this context. This study significantly differs from previously published similar studies in the literature. The research method consists of the SLR proposed by Kitchenham and Charters (2007) and bibliometric analysis. Our search shows that these two research methods have not previously been combined in this format.

The following section briefly discusses lean production. Next, the details of the SLR and bibliometric analysis are presented in the methods section. Then, the findings of SLR and bibliometric analysis are presented in the discussion section along with the limitations of the study, and future research directions.

LEAN PRODUCTION AND CRITICAL SUCCESS FACTORS

Constantly changing global competition and customer expectations forced manufacturing and service firms to adopt new production and service delivery approaches in the globalized marketplace over the last decades (Meredith and Tavish, 1992; Monden, 1998; Panizzolo, 1998; Pakdil and Leonard, 2014). In this complex context, LP has been implemented as an effective method by a great variety of industries in large-sized organizations, including automotive, aerospace, as well as SMEs.

Lean efforts and applications have been analyzed over the last decades, resulting in various definitions of what lean means. Although most definitions are production-focused, there are a few definitions that focus on an overall management perspective. The word *lean* was introduced by MIT professors to name Toyota's production system (Womack and Jones, 1996; Womack et al., 1990). According to Liker (1998), "Leanness is a process, a journey, not an end state." Wilson (2010) stated, "The lean solution is a large paradigm shift." Lean is also seen as a multifaceted concept by various researchers (Bentley et al., 2000; Womack et al., 1990). Creese (2000) identifies lean as "a manufacturing philosophy to shorten lead time and reduce costs by redirecting waste and improving employee performance, skill, and satisfaction". Lean focuses on finding and removing waste instead of adding resources (Foss et al., 2011). A *lean enterprise* is an organization that uses lean concepts and practices, not just in the manufacturing function, but in everything it does (Womack et al., 1990). However, most definitions concentrate on the point of production in organizations. For example, according to Shah and Ward (2007: 791), "lean production is an integrated socio-technical system whose main objective is to eliminate waste by concurrently reducing or minimizing supplier, customer, and internal variability." Womack and Jones (2003) indicated that lean principles are: specifying the value, identifying the value stream, flow, pull, and perfection. A key strategy to supply value to the customer must be eliminating waste. As illustrated often in the literature, the TPS is seen as one of the most successful applications of the lean concept in the world (Wilson, 2010: 9). LP aims to eliminate seven crucial wastes, namely excessive inventory, overproduction, motion, handling, processing, waiting, and correction of defects (e.g., Liker, 1998; Womack and Jones, 1996). Liker (2004) categorized 14 lean principles into two pillars: 1) continuous improvement (CI) and 2) respect for people. These 14 principles are also grouped into four groups: 1) philosophy – long-term, 2) process – promote flow, 3) people and partners– respect and development, and 4) problem solving – CI.

Rockart (1979) defines CSFs as "areas in which results, if they are satisfactory, will ensure successful competitive performance for the organization." They are critical areas for planning and action necessary to obtain effectiveness. They are internal factors that an organization controls; however, external factors are left out since they cannot be controlled through implementation. In regard to SMEs, Achanga et al. (2006) studied four top-level CSFs that may determine whether a lean project is successful or not: leadership and management, financial capabilities, skills and expertise, and organizational culture. However, they argue that these four can be broken down into more components.

However, according to Bhamu and Sangwan (2014), the adoption of lean manufacturing has been limited by SMEs for the implementation costs and the automotive industry has been the focus of research. Yet, with Industry 4.0 the interest in lean is expected to increase as they complement each other by focusing on operational excellence (Moeuf et al., 2018) and supporting environmentally sustainable manufacturing (de Sousa Jabbour et al., 2018). Yet lack of expertise and leadership seems to be a challenge for SMEs (Achanga et al., 2006; Moeuf et al., 2016). Finally, it is vital to keep in mind that CSFs are not the same for every country; although some CSFs overlap for some countries, such as management commitment, education and training, and cultural change (Stankalla et al., 2018).

METHODS

The methods of this study include two main tools: SLR and bibliometric analysis. The SLR method proposed by Kitchenham and Charters (2007) was implemented in this study. The SLR process in this study consisted of three main stages: 1) planning, 2) conducting the review, and 3) reporting the review, as detailed in Figure 1. The following subsections present step-by-step details of the SLR and bibliometric analysis.

Planning Stage

Identify the need for a systematic literature review

The SLR provides researchers with a well-structured method to search the relevant literature in depth (Kitchenham and Charters, 2007). Additionally, the SLR also aids researchers in determining current approaches implemented in previous studies and deciding what approaches could be helpful in the following studies. In this regard, the outcomes of the SLR have the potential to function as an overarching guide in shaping future studies in the discipline. With this in mind, the need identified in this SLR is to identify the CSFs and barriers of LP, particularly in SMEs.

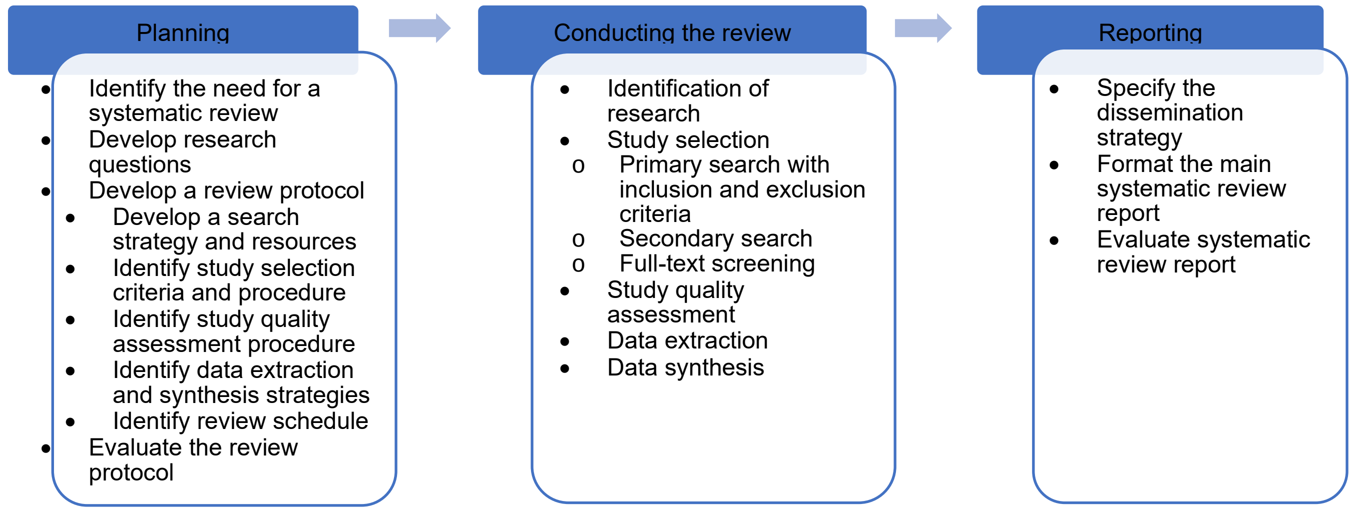


Figure 1: The SLR process

<u>Develop research questions</u>

Kitchenham and Charters (2007: 9) assert that "specifying the research questions is the most important part of any systematic review." Given the need identified for a systematic review as above, this study was constructed upon the following research question:

RQ1: What are the CSFs and barriers that affect the performance of LP in the SMEs?

<u>Develop a review protocol</u>

Kitchenham and Charters (2007) state that using a well-established review protocol minimizes the potential bias of the researchers. A review protocol in SLR comprises several elements: i) develop a search strategy and identify resources, ii) identify study selection criteria and procedure, iii) identify study quality assessment procedure, iv) identify data extraction and synthesis strategies, and v) identify review schedule.

While developing a search strategy and identifying resources, digital databases were identified for the primary and secondary searches in the SLR. The primary search included EBSCO, Web of Science, and Scopus digital databases since these were considered the most comprehensive digital databases that cover the most relevant publications. A secondary search was performed, followed by the primary search to consider the articles cited in the references of the articles identified in the primary search.

Study selection criteria is set up to assess the articles linked with the review's research question (Kitchenham and Charters, 2007). Study selection criteria include keywords and inclusion/exclusion criteria of the research. The keywords used in the SLR are 1) lean production, 2) critical success factor, and 3) SMEs. Keywords were searched only in "title" and "abstract" and constructed using Boolean ANDs. Inclusion/exclusion criteria were determined to ensure that all relevant previous studies were added to answer the research question of the SLR (Table 1).

Table 1: Inclusion/exclusion criteria

Criteria	Items
Inclusion criteria	• Full-text articles published in scholarly journals, theses, book chapters, and books • Published in English between 2000 and 2021 • The publications found in selected digital databases in the primary search • The articles including identified keywords • The articles used in the previous literature reviews (secondary search)
Exclusion criteria	• Articles published in non-scholarly journals • Not full-text articles • Materials published in other languages • Duplicated studies • Conference proceedings

There is no consensus on the definition of study quality in SLR (Kitchenham and Charters, 2007. Cochrane Reviewers' Handbook (2003) reports that the quality study is associated with minimizing bias and internal/external validity of the study. To perform a quality assessment in the SLR, the following questions were answered in the quality assessment step:

QA1: Was the identified digital database capable of covering potential articles published about keywords and inclusion/exclusion criteria?

QA2: Were the inclusion/exclusion criteria thoroughly followed in the digital database in this SLR?

QA3: Was the topic of each article directly related to the keywords used in this study?

QA4: Was the method employed in the article explicitly and clearly detailed and described?

For the data extraction and synthesis methods step, the researchers implemented full-text scanning, where each article was read and analyzed in depth. To identify the review schedule, a timetable that includes the steps of the SLR was constructed by the researchers.

Evaluate the review protocol

The review protocol constructed in the previous step was assessed based on resource availability and feasibility. The researchers aimed to use the resources effectively and efficiently throughout the SLR process. Evaluation outcomes revealed that the time devoted to this research was enough to complete it on time. Additionally, resources provided by the universities, such as digital databases, were present and available for this SLR.

Conducting the review stage

Identification of research

This step heavily focused on the research question and review protocol created in this study. The outcomes were listed in multiple matrix tables created in MS Excel that include the title, year, authors, volume, issue, page numbers, abstracts of the publications, and the names of the journals. The tables were analyzed and filtered by several variables.

Study selection

The primary search found 229 articles that match with the keywords and inclusion/exclusion criteria. Each article was read and analyzed in the full-text scanning process to exclude irrelevant articles, and 54 articles were discarded. This process left a total of 175 articles. The secondary search included relevant articles cited in the articles analyzed in the full-text screening process and additional 20 articles that were not listed in the primary search. At the end of primary and secondary searches, SLR resulted in a total of 78 articles.

Study quality assessment

For QA1, the number of articles found in the digital database was deemed enough to search the relevant literature. For the following three questions, each article has been reviewed using keywords and inclusion/exclusion criteria in full text screening. It was found that 116 irrelevant articles were already excluded in this process. The researchers ensured that the inclusion/exclusion criteria were thoroughly applied in SLR.

Data extraction

The matrix tables were used to extract data through full-text screening. While extracting data, each article was listed with a unique number along with the following variables: 1) article information, including author/s, title, year, and the journal name, 2) abstract, 3) CSFs, 4) barriers, and 5) an assessment score assigned by the researchers to identify the degree of the relevancy of the article to the research question. Similar CSFs identified in the reviewed articles were listed under one unique factor (Appendix 1). A similar consolidation process was run to analyze the barriers in the reviewed articles (Appendix 2).

Reporting the review stage

The data extracted in the previous steps were synthesized to answer the research question of the SLR at this stage. As a reminder, this study aimed to identify the CSFs and barriers of LP systems implemented in SMEs. The findings are presented in Section 4.

Bibliometric method

Bibliometrics is the utilization of statistical and quantitative techniques on scholarly communication (Pritchard, 1969) and helps us understand the cognitive structure of a field or concept (Garfield, 1972; Moed, 2005). Bibliometrics is widely used for summarizing the most representative results of a set of bibliographic documents. Many fields and concepts are analyzed using bibliometric methods. Molecular biology (He and Teng, 2005), industrial engineering (Cancino et al., 2017), management (Podsakoff et al., 2008), knowledge management (Gu, 2004), economics (Bonilla et al., 2015), innovation (Fagerberg et al., 2012), entrepreneurship (Landstrom et al., 2012), origins of life research (Aydinoglu and Taskin, 2018), quantum information technologies (Seskir and Aydinoglu, 2021) are to name a few of these studies.

Furthermore, recently bibliometric methods and SLR have been employed complementarily. The former helps the researchers to overview bigger datasets via statistical and quantitative techniques, while the latter allows a qualitative approach to assess the relevant scholarly literature in detail. The macro and micro perspectives are obtained together; hence, it is considered a faster and more effective method (Pulsiri and Vatananan-Thesenvitz, 2018). Pattnaik, Kumar, and Vashista (2020) investigated trade credit system literature employing both techniques to identify research gaps; Janowski (2020) looked at corporate social responsibility literature through the contributions of Andrew Carnegie's philanthropic activities, and Cardella, Hernández-Sánchez and Sanchez Garcia (2020) studied the relationship between family role and entrepreneurship from a multidisciplinary perspective. The combination of methods is used even when there are smaller datasets, such as the study of the impact of COVID-19 on entrepreneurial ventures, which introduced the concept of entrepreneurial interventions (Krishnan, Ganesh, and Rajendran, 2022). In a nutshell, SLR and bibliometric methods are powerful tools for evaluating a body of literature. To our knowledge, this is the first study to do that for CSFs and lean production in SMEs.

There are several scholarly databases, such as Web of Science, Scopus, and Dimensions, used in the bibliometric analysis. We used Scopus, which has the most extensive title coverage that includes more than 23,000 active, high-quality journals and provides enough information about the publications, such as authors' names, affiliations, abstract, keywords, and subject areas. This information allowed us to extract high-value analyzes that could not be performed using other databases.

For creating the dataset, we searched in the topic, abstract or keywords sections the research query: ("lean production" OR "critical success factor") AND ("sme" OR "smes" OR "small and medium enterprise" OR "small to medium enterprise" OR "small-and-medium enterprise"). As the bibliometric databases do not necessarily have full access to the articles, these fields are created for ease of finding relevant research. They significantly impact on discoverability (Bekhuis, 2015; Wang and Mi, 2012). Our search returned with 562 publications on January 8, 2022.

We used VosViewer for network analysis (VosViewer, 2021; van Eck and Waltman, 2013) and bibliometrix package of R statistics program (R Bibliometrix, 2021). We downloaded the data in *tabdelimited* format and curated it for our analysis.

RESULTS

As stated in the research question, the results were categorized into two folds: CSFs and barriers of LP systems in SMEs. First, the CSFs were analyzed in 78 articles and found in 68 articles. These factors with the frequencies and citing articles are presented in Appendix 1. The frequencies and percentages of the CSFs help SMEs identify which factors would be more influential in lean efforts. Appendix 1 demonstrates that 61 unique factors were identified in the SLR. Among these factors, “management commitment” is revealed as the most frequent CSF listed by 32 articles (8.7%). Next, leadership, employee commitment and engagement, knowledge of lean tools, communication, project management, training, strategic management, change management, effective HRM, organizational culture, and resource management are listed as the most frequent twelve CSFs.

Second, the barriers identified in the literature were analyzed in 78 articles and found in 24 articles. Appendix 2 demonstrates the frequencies and percentages of the barriers found in the reviewed articles. The frequencies of the barriers shed some light on the lean efforts implemented in SMEs. These barriers assist decision-makers with developing effective lean implementation plans and starting the lean journey to prepare for potential challenges. The barriers also enable lean practitioners to understand why lean efforts would fail in SMEs. These points demonstrate what should not be done and what potential barriers would appear in the lean journey. “Resistance to change" appears to be the most frequent barrier in the literature. This finding is consistent with the literature and practice, and resistance to change is known as one of the critical research areas in organizational behavior discipline. Extensive research shows that lean transformation fails if the lean implementation does not consider the degree of resistance to change across the organization (e.g., Sim and Chiang, 2012; Azyan et al., 2017). Given that less than 30% of the organizations succeeded in lean implementation over the last decades (Bhasin and Burcher, 2006; Jadhav et al., 2014; Camagu, 2010; Mohanty et al., 2007; Sing et al., 2010), this finding points out that lean implementation should prioritize “resistance to change" as a potential barrier for lean success and that additional plans should be integrated into lean implementation for overcoming the consequences of this particular barrier. That finding also confirms that lean transformation in business organizations requires a systematic paradigm change and cultural transformation prior to effective lean implementation (e.g., Womack et al., 1990; Pakdil and Leonard, 2017). If the organization faces serious resistance to that change, the entire lean implementation is expected to identify the root causes of that resistance before beginning an effective lean implementation.

The second barrier that has the highest frequency is “lack of training and knowledge.” Proper training and knowledge of lean is an essential requirement for successful lean implementation (Secchi and Camuffo, 2016; Womack et al., 1990; Monden, 1998). “Institute training on the job”, which is among Deming’s 14 points, should be considered while planning the detailed action plans of the lean implementation. The lack of training and knowledge of lean is likely to cause resistance to change and loss of employee commitment in the long run (e.g., Fournier et al., 2021). Similar to CSFs, “lack of leadership commitment and support” was also listed as a barrier in the reviewed studies. That finding reflects the pivotal role of leadership in lean management. Additionally, “lack of resources” was listed as the fourth significant barrier in the reviewed studies. The fact that SMEs do not have access to resources that large-sized organizations may easily have (Netland, 2016) makes that barrier more critical for the lean success of SMEs.

In addition to analyzing the CSFs and barriers, a few additional variables were considered for further analysis. The journals that published the reviewed articles were analyzed as presented in Figure 2. The analysis shows that the *International Journal of Production Research* is the leading journal publishing most articles. The mission of the journal aligns with the pivotal role of CSFs and barriers of lean. It is followed by the *International Journal of Advanced Manufacturing Technology* with four studies.

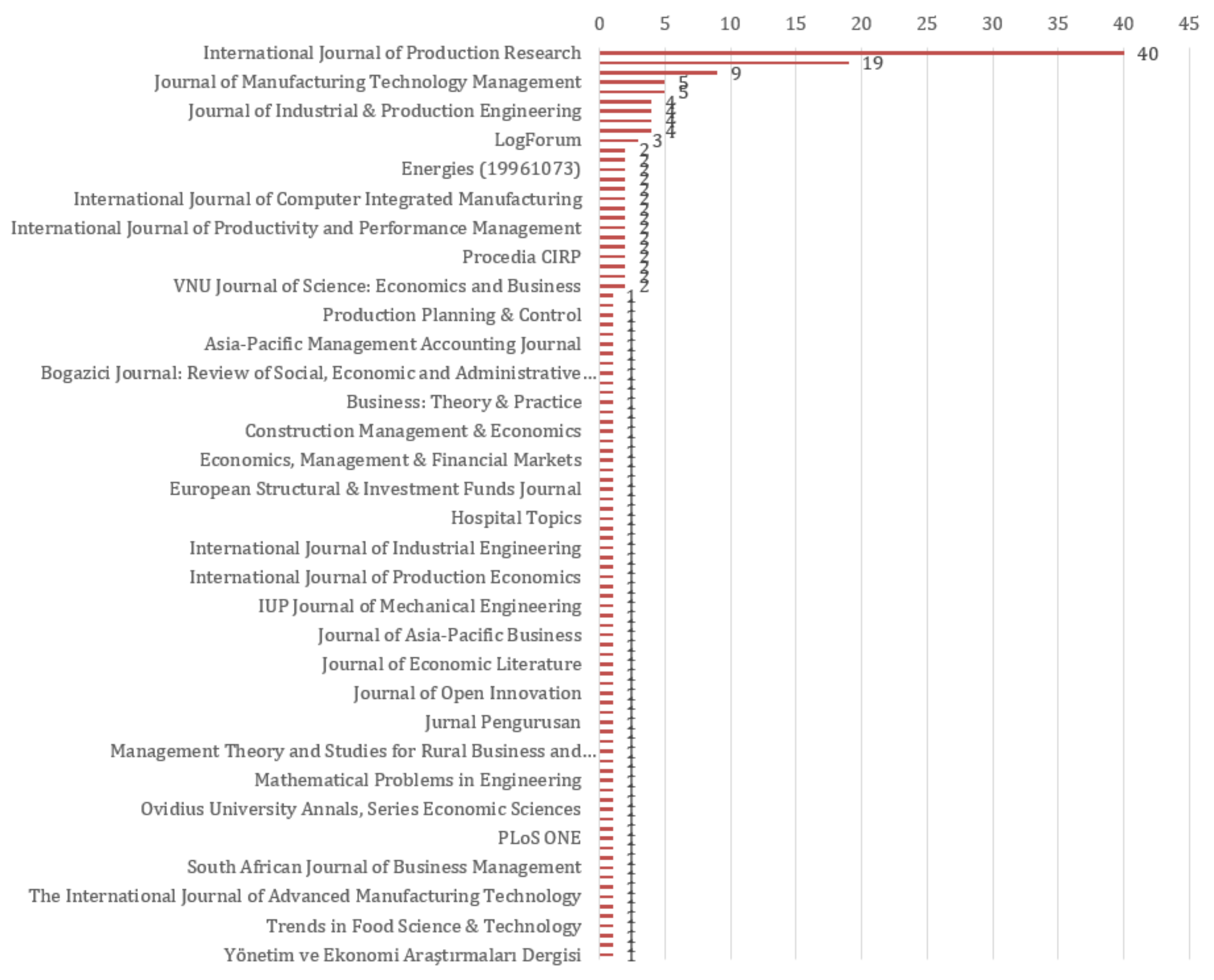


Figure 2. Publications by the journals

Around 60% of the 562 publications used in the bibliometric analysis are research articles; the rest are conference papers, book chapters, and review papers. As seen in Figure 3, there has been a growing scholarly interest in LP and CSFs in SMEs since the 2010s.

When we check the subject areas (Scopus in-house experts identify subject areas based on the aims and scope of the title and on the content it publishes) of these publications, the multidisciplinarity of the LP becomes apparent. Slightly over a quarter of publications are published in Business, Management, and Accounting journals, whereas slightly less than a quarter are published in engineering journals. Computer science journals (18.1%) and decision sciences journals (12.0%) are the other notable fields that host similar research. The rest are less than 4% but somehow demonstrate the breadth of the topic. Social Sciences; Economics, Econometrics, Finance; Mathematics; and Material Sciences are to name a few (Figure 4). Therefore, we can argue that LP sits at the nexus of a number of research fields contributing to knowledge transfer among them.

In order to understand the most critical topics, we checked the most frequent 20 keywords and phrases in the publications. SMEs (267), CSFs (163), LP (147), and enterprise resource planning (110) are the most frequent concepts. They are followed by manufacture, agile manufacturing system, survey, competition, industry, and so forth (Figure 5).

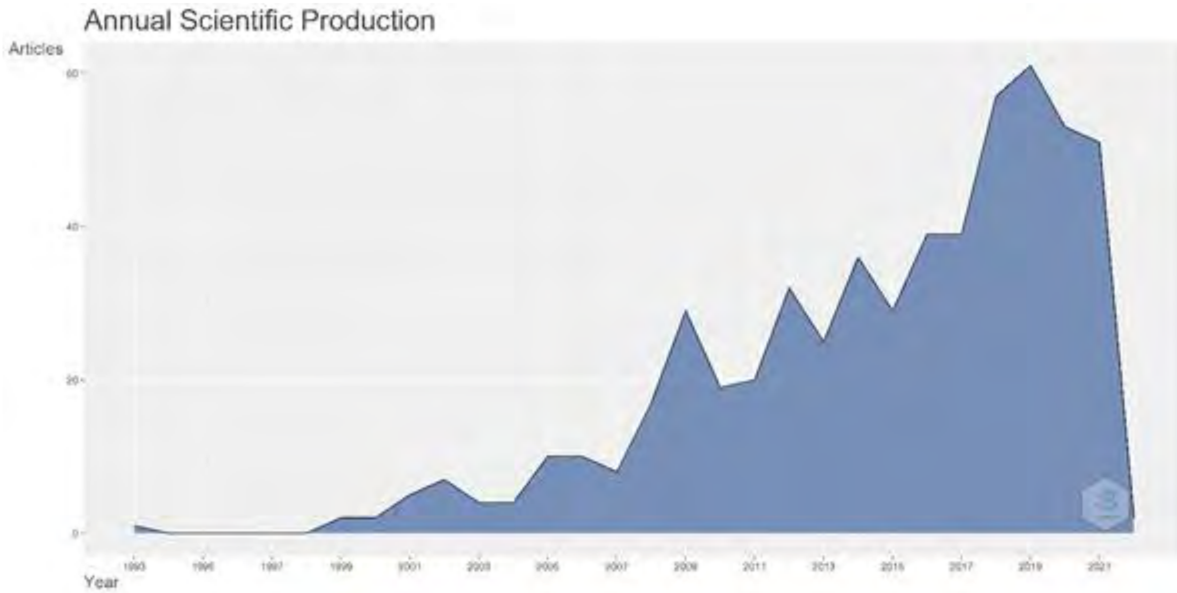


Figure 3: Number of publications per year

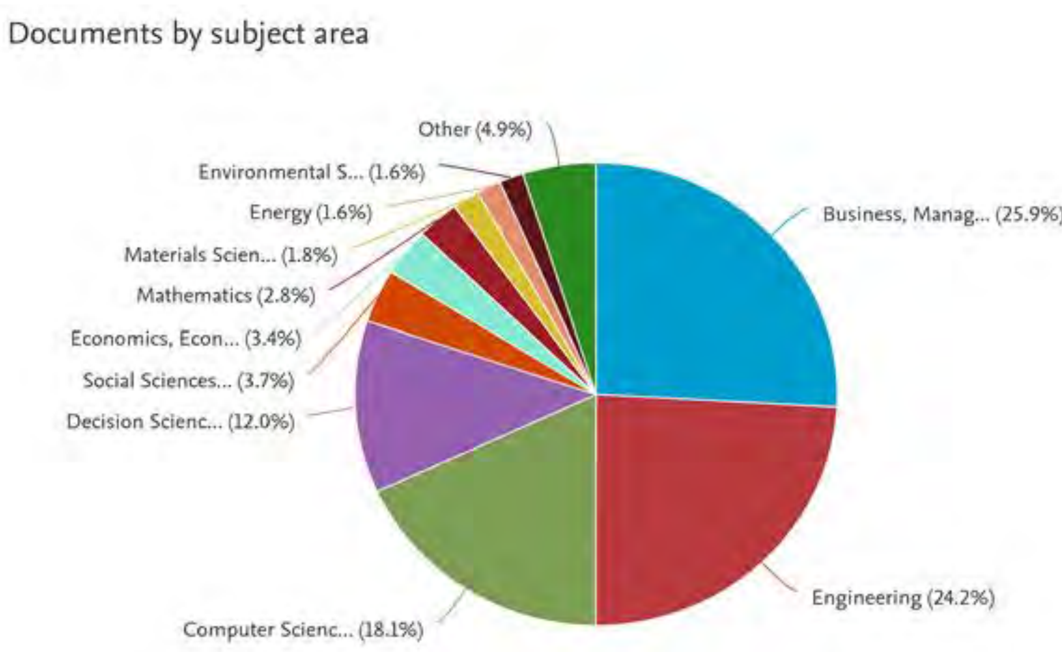


Figure 4: Publications by subject area

We applied a keyword analysis algorithm that identifies color-coded clusters and the relationship among them (van Eck and Waltman, 2013). The algorithm analyses the words or phrases in the titles, abstracts, and keywords and puts them in the same cluster if they are related. The size of the nodes corresponds to the frequency of the term or phrase in our dataset. The lines between nodes mean that these two concepts appear in the same section, such as the same sentence, which indicates a cognitive relationship. The distance between the nodes also indicates a closer relationship in the visual below.

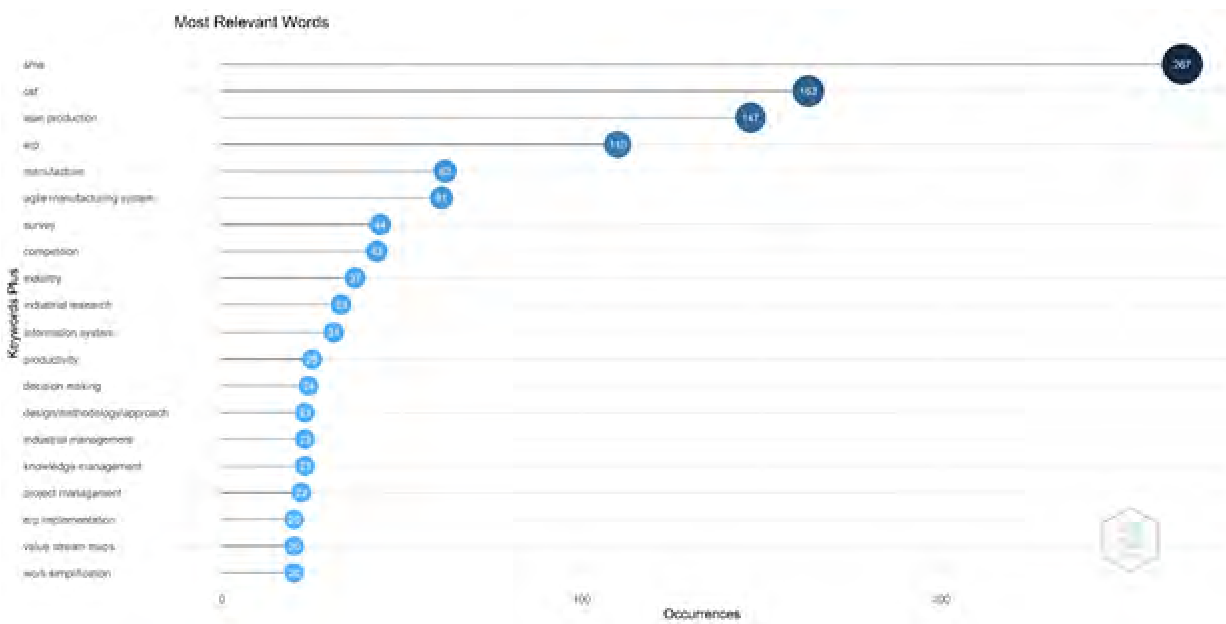


Figure 5: Most frequent concepts in LP and CSFs related publications

We identified three major and four minor clusters in Figure 6. The biggest cluster (red) is about SMEs and CSFs. Other notable nodes in Figure 6 are enterprise resource planning (ERP), industry, knowledge management, resource allocation, information use, cloud computing, strategy, management science, and research. This cluster seems to focus on CSFs; however, SMEs have a central position and connect many nodes. The second cluster (green) has more nodes than the third one (blue) and is better spread into the network. The nodes, such as manufacture, lean implementation, industrial research, design/methodology/approach, lean practices, industrial management, performance, and LP system, indicate that the green cluster concepts deal with the “how-to” of the lean concept. The third node (blue) is about LP and manufacturing processes. LP, lean manufacturing, agile manufacturing system, value stream maps, productivity, CI/production management, and kaizen are among the biggest nodes in the cluster. All these three clusters exhibit an intertwined relationship with the rest of the network. However, the fourth cluster (yellow), mostly favors relationships with the red cluster. The nodes include competition, decision making, process engineering, SS, lean SS, benchmarking, and decision support system. The rest of the clusters and nodes are separated in the network.

Finally, we applied a thematic evolution algorithm to conduct a temporal analysis of the key concepts regarding LP and CSFs (Figure 7). The analysis provides us with conceptual subdomains that would further our understanding as follows: (1) Themes in the upper-right quadrant are both well developed and represent the main foundations of a research field. They are known as the *motor themes* of the specialty, given that they present strong centrality and high density. The placement of themes in this quadrant implies that they are related externally to concepts applicable to other themes that are conceptually closely related. Figure 8 shows that “LP”, “manufacture”, and “agile manufacturing system” are listed in motor themes in this study. (2) Themes in the upper-left quadrant have well-developed internal ties, but unimportant external ties and are of only marginal importance for the field because they are very specialized or peripheral. These themes are called *niche theme*s. Here we see “Malaysia”, “knowledge-based system”, and “software process improvement” as the three niche themes. There are definitely not mainstream topics for CSF in LP, but there are a number of studies. (3) The themes in the lower-left quadrant, called *emerging or declining themes*, are weakly developed and marginal. The themes of this quadrant have low density and low centrality, mainly representing either emerging or disappearing themes. The concepts “supply chain”, “factor analysis”, and “supply chain management” appearing in this quadrant can be interpreted as emerging themes since LP heavily relies on inventory reduction, zero inventory, and supply chain optimization. This finding is supported by the SLR component of this study as a number of studies identified supply chain related concepts (Dehdasht, et al., 2020; Narkhede, et al., 2020). We believe this is an emerging topic, not a declining one. (4) The themes in the lower-right quadrant, *basic themes,* are essential for a research field but have a low degree of development. Therefore, this quadrant groups transversal and general, basic themes” (Cobo et al., 2011).

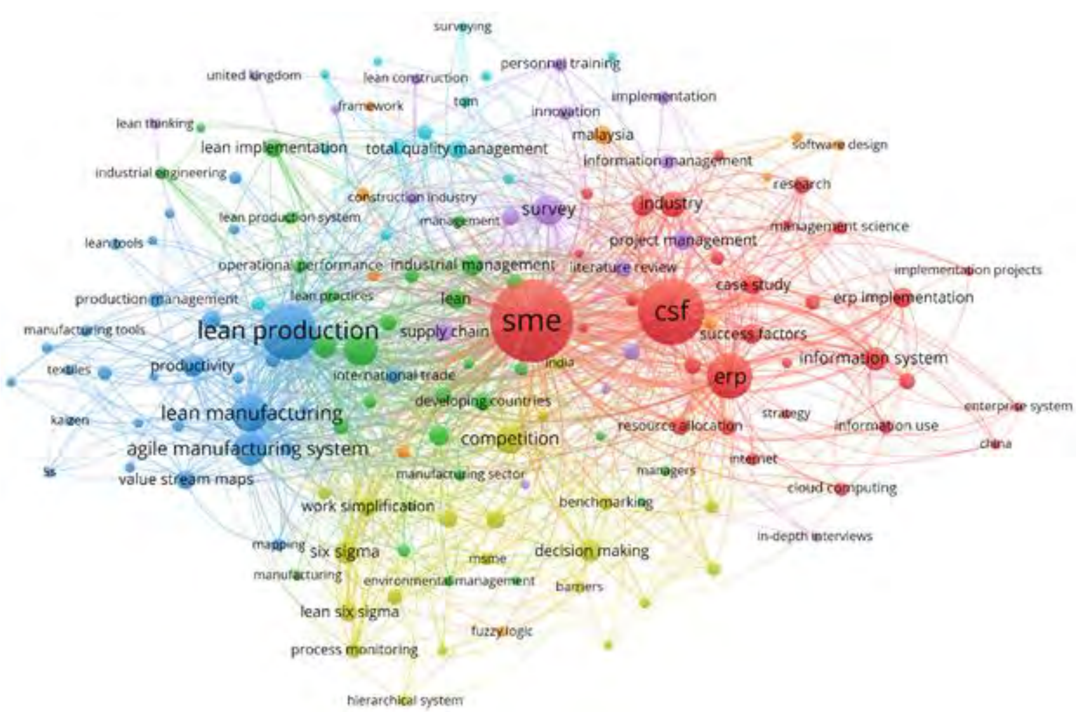


Figure 6: The clusters generated in the bibliometric analysis

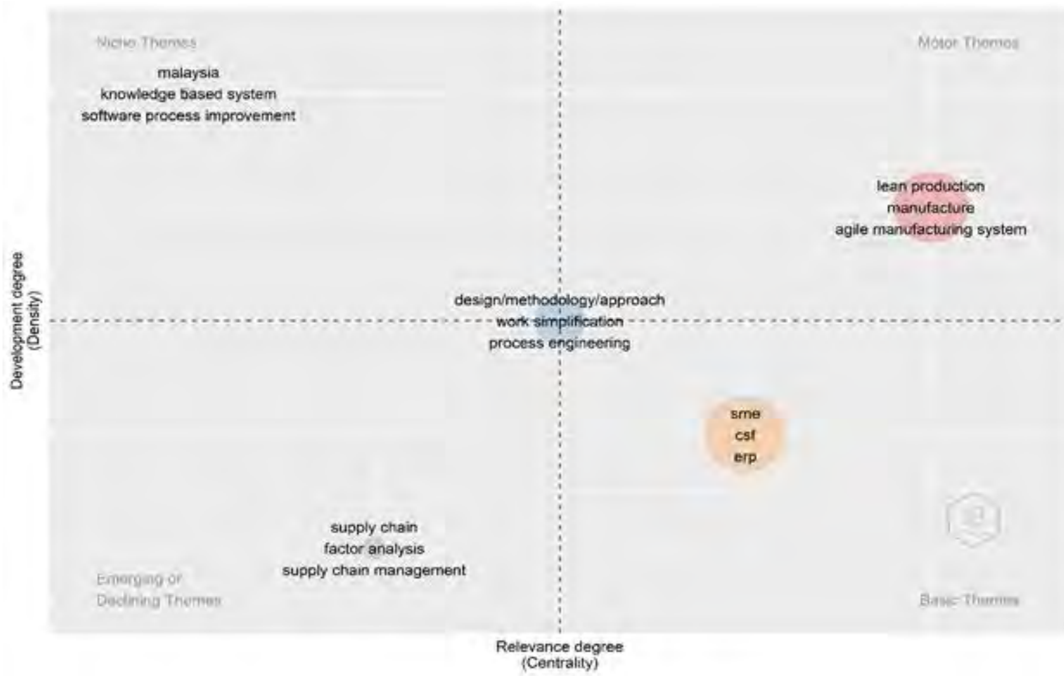


Figure 7: Thematic network for LP and CSFs related publications

DISCUSSION

The results of this study show that certain CSFs are paramount for successful lean implementation. Management commitment, leadership, employee commitment and engagement, and knowledge of lean tools are the most frequently used factors in the previous literature. As Deming's 14 points (Deming, 1986) pointed out, top management responsibility and leadership are still seen as the two critical requirements for lean success. Literature and practice point out that top management is responsible for assuring the success of implementing a new management philosophy and is perceived as a role model in lean (Kanning and Bergmann, 2009; Zargun et al., 2014; Alefari et al., 2017). Literature also requires lean leadership (Liker and Convis, 2012; Dombrowski and Mielke, 2014) as an effective concept that contains specific leadership behaviors and competencies that promote lean implementation by focusing on particular practices such as employee empowerment, training, and teamwork within the organization (Poksinksa et al., 2013; Bodek 2008). Additionally, Emiliani and Stec (2005) state that if senior managers fail to understand the complex structure of lean systems, lean transformation may result in less desirable outcomes. In this context, it is evident that lean implementation requires strong support and commitment and effective lean leadership from the top management (Shah and Ward, 2007; Hayes et al., 2005; Mann et al., 2009, Achanga et al., 2006; Moeuf et al., 2016).

"Employee commitment and engagement" has been the third most frequently used CSF in the reviewed literature. Employee engagement is defined as the extent of commitment and involvement an employee has towards the organization and its values (Anitha, 2014: 308). The antecedents of employee engagement are listed as job factors, perceived organizational support, rewards, recognition, procedural and distributive justice (Saks, 2006), and organizational culture (Compton and Granito, 2002). Employee engagement concept has been presented as a strong predictor of increasing organizational performance (Markos and Sridevi, 2010). In a similar vein, employee commitment and engagement and “respect for people” concepts are presented as the backbones of lean success to sustain continuous quality improvement (Reid and Sanders, 2021; Womack et al., 1990).

“Knowledge of lean tools” has been identified as the fourth most frequently used CSF in this study. Learning and understanding lean tools are essential for successful lean implementation and lean sustainability (Zhang et al., 2020). Knowing how to use lean tools appropriately improves the success of lean implementation. In other words, successful lean implementation requires a certain level of knowledge of lean and needs effective and continuous training programs for the employees. Zhang et al. (2020) assert that using lean tools is a process of knowledge creation that contains knowledge transformation and use in lean practices. Literature indicates that a lack of knowledge of lean tools is likely to fail in lean implementation and practices (Harris et al., 2014; Popoola, 2000). lack of knowledge of lean tools and such failure may also cause a lack of belief in the benefits of lean practices. To improve the current knowledge level of lean tools, continuous lean training and education functions are implemented and sustained in the organizations (Womack et al., 1990; Harris et al., 2014; Pakdil and Leonard, 2014).

Our bibliometrics findings support the SLR as top management support, knowledge management, personnel training, lean tools, resource allocation, and information use are among the keywords of the publications. However, commitment and engagement are not detected. This finding is not surprising as bibliometric methods are not “smart”; they match the words but cannot extrapolate the meanings. For instance, resource allocation can be considered a proxy for management commitment. Furthermore, it is not expected to have a 100% match between SLR and bibliometrics. That being said, the similarity of the findings of the separate datasets (EBSCO, Web of Science, and Scopus vs only Scopus) and methodologies (SLR vs bibliometrics) can be considered a confirmation.

The results of this study also show a great variety of barriers identified in the relevant literature and practices. The first four barriers are resistance to change, lack of training and knowledge, lack of leadership commitment and support, and lack of resources. Research shows that when the employees receive adequate, timely, and helpful information about the changes and are involved in decision-making and change processes, the resistance to change is likely to decrease in lean transformation (Tran et al., 2020; Canning and Found, 2015). Additionally, when the employees are trained on tools and knowledge about the change, the degree of resistance to change significantly decreases in lean transformation (Tran et al., 2020). Similarly, if senior leadership participation and communication are provided in lean transformation, resistance to change is minimized (Burmester, 2017). Research also shows that industry-specific resistance to change in lean, such as physicians’ and higher education employees’ resistance to change, should also be considered prior to starting lean transformation (Fournier et al., 2021; Allaoui and Benmoussa, 2019).

Lack of resources is identified as one of the critical barriers in lean implementation in the literature regardless of the size and industry of the organizations. Top management is ultimately responsible for providing adequate resources on time for lean success (e.g., Martinez-Jurado and Moyano-Fuentes, 2012). Limited financial, managerial, physical, and operational resources may also result in increased resistance to change as a barrier in lean (Albliwi et al., 2014). As a critical argument, all necessary resources should be provided by the organizations before starting the lean implementation to overcome the potential lack of resources-related barriers and obstacles.

As a conclusion, this study focuses on exploring the CSFs and barriers of LP in SMEs. After performing an SLR, "management commitment" was found to be the most frequent CSF, and "resistance to change" was the most frequent barrier to lean among SMEs in the reviewed articles. The details of the findings of this study are expected to be insightful for the SMEs that plan to start lean journeys. As well as the strengths, this study may have some limitations. The databases and timeline may be one of the potential limitations. Using other databases and a different timeline may have included some other papers in the SLR. The other limitations may be related to the steps followed in the SLR. For example, if the inclusion/exclusion criteria are set up differently, this study's results may also vary. To eliminate this, future research may focus on a broader timeline and other databases that were not included in this study.

REFERENCES

Abolhassani, A. & Gopalakrishnan, B. (2016). Lean and US manufacturing industry: popularity of practices and implementation barriers. *International Journal of Productivity and Performance Management,* 65(7): 875-897.

Achanga, P., Shehab, E., Roy, R. & Nelder, G. (2006). Critical success factors for lean implementation within SMEs. *Journal of Manufacturing Technology Management,* 17(4): 460-471.

AP, A., EK, A., Nm, I., & Ee, I. (2020). Lean management and performance of telecommunication firms: The mediating role of workplace structure. *Cogent Business & Management*, *7*(1): 1761637.

Alefari, M., Salonitis, K. & Xu, Y. (2017). The role of leadership in implementing lean manufacturing. *Procedia Cirp,* 63: 756-761.

Alhuraish, İ., Robledo, C. & Kobi, A. (2017). A comparative exploration of lean manufacturing and six sigma in terms of their critical success factors. *Journal of Cleaner Production,* 164: 325-337.

Alibhai, S., Bell, S. & Conner, G. (2017). *What's Happening in the Missing Middle?: Lessons from Financing SMEs,* Washington, DC.: World Bank. https://openknowledge.worldbank.org/handle/10986/26324 License: CC BY 3.0 IGO.

Allaoui, A., & Benmoussa, R. (2020). Employees' attitudes toward change with Lean Higher Education in Moroccan public universities. *Journal of Organizational Change Management, 33*(2): 253-288.

Alsmadi, M., Lehaney, B. & Khan, Z. (2012). Implementing Six Sigma in Saudi Arabia: An empirical study on the fortune 100 firms. *Total Quality Management and Business Excellence.* 23(3/4): 263-276.

Andreadis, E., Garza-Reyes, J. A. & Kumar, V. (2017). Towards a conceptual framework for value stream mapping (VSM) implementation: an investigation of managerial factors. *International Journal of Production Research,* 55(23): 7073-7095.

Anholon, R. & Sano, A. (2016). Analysis of critical processes in the implementation of lean manufacturing projects using project management guidelines." *International Journal of Advanced Manufacturing Technology,* 84(9-12): 2247-2256.

Antony, J. & Desai, D. A. (2009). Assessing the status of six sigma implementation in the Indian industry: Results from an exploratory empirical study. *Management Research News,* 32(5).

Antony, J. & Kumar, M. (2012). Lean and Six Sigma Methodologies In NHS Scotland: An Empirical Study And Directions For Future Research. *Quality Innovation Prosperity / Kvalita Inovácia Prosperita,* 16(2): 19-34.

Antony, J., Kumar, M. & Labib, A. (2008). Gearing Six Sigma into UK manufacturing SMEs: results from a pilot study. *Journal of the Operational Research Society,* 59(4): 482-493.

Arabi, S., Chafi, A. & El Hammoumi, M. (2021). Bibliographic Study on the Difficulties Encountered by SMEs. *Journal of Advanced Manufacturing Systems,* 20(01): 163-190.

Ayçin, E., (2021). Yalın Üretim Uygulamalarında Karşılaşılan Engellerin Gri Dematel Bütünleşik Yaklaşımıyla Analizi. *Yönetim ve Ekonomi Araştırmaları Dergisi,* 17(3): 1-18.

Aydinoglu, A. U. & Taskin, Z. (2018). Origins of life research: A bibliometric approach. *Origins of Life and Evolutions of Biospheres,* 48(1): 55-71.

Azyan, Z., Pulakanam, V. & Pons, D. (2017). Success factors and barriers to implementing lean in the printing industry: a case study and theoretical framework. *Journal of Manufacturing Technology Management,* 28(4).

Badgujar, P., Kanungo, B. & Thakar, G. D. (2016). Identification of Factors Affecting Lean Manufacturing Implementation in Pump Manufacturing Companies in India A Case Study. *International Journal for Quality Research,* 10(3): 495-510.

Bagni, G., Godinho Filho, M., Luiz Tortorella, G. and Chiappetta Jabbour, C. J. (2021). Critical soft factors for card-based production systems implementation: a multi-method study. *Brazilian Journal of Operations & Production Management,* 18(4): 1-17.

Baskaran, S. M. & Lakshmanan, A. R. (2019). A Framework for Lean Readiness Evaluation Using a Hierarchical Fuzzy System. *South African Journal of Industrial Engineering,* 30(1): 220-234.

Bekhuis, T. (2015). Keywords, discoverability, and impact. *Journal of the Medical Library Association: JMLA*, *103*(3): 119.

Belhadi, A., Touriki, F. E. & El Fezazi, S. (2016). A Framework for Effective Implementation of Lean Production in Small and Medium-sized Enterprises. *Journal of Industrial Engineering and Management.,* 9(3): 786-810.

Belhadi, A., Sha'ri, Y. B. M., Touriki, F. E. & El Fezazi, S. (2018). Lean production in SMEs: literature review and reflection on future challenges. *Journal of Industrial and Production Engineering,* 35(6): 368-382.

Belhadi, A., Touriki, F. E., & Elfezazi, S. (2019). Evaluation of critical success factors (CSFs) to lean implementation in SMEs using AHP: A case study. *International Journal of Lean Six Sigma*, *10*(3): 803-829.

Bentley, G., Nightingale, G. & Taneja, A. (2000). *Lean Enterprise Model,* s.l.: http://lean.mit.edu/events/workshops/files_public/Le_lem.pdf.

Berlec, T., Kleindienst, M., Rabitsch, C. & Ramsauer, C. (2017). Methodology to Facilitate Successful Lean Implementation. *Journal of Mechanical Engineering,* 63(7/8): 457-465.

Bevilacqua, M., Ciarapica, F. E. & De Sanctis, I. (2017). Lean practices implementation and their relationships with operational responsiveness and company performance: an Italian study. *International Journal of Production Research,* 55(3): 769-794.

Bhadu, J., Bhamu, J., & Singh, D. (2021). Critical success factors of lean manufacturing for Indian ceramic industries: Modelling through interpretive ranking process. *Indian Journal of Engineering and Materials Sciences*, 28(3): 271-281.

Bhamu, J., & Sangwan, K. S. (2014). Lean manufacturing: literature review and research issues. *International Journal of Operations & Production Management,* 34(7): 876-940.

Bhasin, S. (2012). Prominent obstacles to lean. *International Journal of Productivity and Performance Management,* 61(4): 403-425.

Bhasin, S. & Burcher, P. (2006). Lean viewed as a philosophy. *Journal of Manufacturing Technology Management,* 17(1): 56-72.

Binti Aminuddin, N. A., Garza-Reyes, J. A., Kumar, V., Antony, J., and Rocha-Lona, L. (2016). An analysis of managerial factors affecting the implementation and use of overall equipment effectiveness. *International Journal of Production Research*, *54*(15): 4430-4447.

Bodek, N. (2008). Leadership is critical to Lean. *Manufacturing Engineering,* 140(3): 145-154.

Bonilla, C., Merigó, J. M. & Torres-Abad, C. (2015). Economics in Latin America: a bibliometric analysis. *Scientometrics,* 105(2): 1239-1252.

Brun, A. (2011). Critical success factors of Six Sigma implementations in Italian companies. *International Journal of Production Economics,* 131(1): 158-164.

Burmester, E. (2017). *Qualitative examination of strategies to overcome resistance to change in lean manufacturing,* Doctoral dissertation, Walden University.

Camagu, S. (2010). *Investigating factors that negatively influence lean implementation in the eastern cape automotive industry.* Master's thesis, Faculty of Business and Economic Sciences, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.

Cancino, C., Merigó, J. M., Coronado, F., Dessouky, Y., & Dessouky, M. (2017). Forty years of Computers & Industrial Engineering: A bibliometric analysis. *Computers & Industrial Engineering*, 113: 614-629.

Canning, J., & Found, P. A. (2015). The effect of resistance in organizational change programmes: A study of a lean transformation. *International Journal of Quality and Service Sciences, 7(*3): 274-295.

Cardella, G. M., Hernández-Sánchez, B. R., & Sanchez Garcia, J. C. (2020). Entrepreneurship and family role: a systematic review of a growing research. *Frontiers in Psychology*, 10: 2939.

Cil, İ. & Turkan, Y. (2013). An ANP-based assessment model for lean enterprise transformation. *International Journal of Advanced Manufacturing Technology,* 64(5-8): 1113-1130.

Cobo, M. J., López-Herrera, A. G. & Herrera-Viedma, E. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics,* 5(1): 146-166.

Compton, D., & Granito, J. (2002). *Managing Fire and Rescue Services*. Washington DC: ICMA University.

Creese, R. C. (2000). *Cost management in lean manufacturing enterprises.* Morgantown, WV., AACE International Transactions, West Virginia University.

Danese, P., Manfè, V. & Romano, P. (2008). A Systematic Literature Review on Recent Lean Research: State-of-the-art and Future Directions. *International Journal of Management Reviews,* 20(2): 579-605.

de Oliveira, R. I., Sousa, S. O. & de Campos, F. C. (2019). Lean manufacturing implementation: bibliometric analysis 2007-2018. *International Journal of Advanced Manufacturing Technology,* 101(1-4): 979-988.

DeSanctis, I., Ordieres Mere, J. B., Bevilacqua, M., & Ciarapica, F. E. (2018). The moderating effects of corporate and national factors on lean projects barriers: a cross-national study. *Production Planning & Control*, *29*(12): 972-991.

de Sousa Jabbour, A. B. L., Jabbour, C. J. C., Foropon, C., & Godinho Filho, M. (2018). When titans meet–Can industry 4.0 revolutionise the environmentally-sustainable manufacturing wave? The role of critical success factors. *Technological Forecasting and Social Change*, 132: 18-25.

Deflorin, P. & Scherrer-Rathje, M. (2012). Challenges in the transformation to lean production from different manufacturing-process choices: a path-dependent perspective. *International Journal of Production Research,* 50(14): 3956-3973.

Dehdasht, G., Ferwati, M. S., Zin, R. M. & Abidin, N. Z. (2020). A hybrid approach using entropy and TOPSIS to select key drivers for a successful and sustainable lean construction implementation. *PLoS ONE,* 15(2): 1-32.

Deming, W. E. (1986). *Out of the Crisis.* Cambridge, MA: MIT Press.

Dolen, T. L. & Hacker, M. E. (2005). A Review of Lean Assessment in Organizations: An Exploratory Study of Lean Practices by Electronics Manufacturers. *Journal of Manufacturing Systems,* 24(1): 55–67.

Dombrowski, U. & Mielke, T. (2013). Lean Leadership – Fundamental Principles and their Application. *Procedia CIRP,* 7: 569-574.

Done, A., Voss, C. & Rytter, N. G. (2011). Best practice interventions: Short-term impact and long-term outcomes. *Journal of Operations Management,* 29(5): 500-513.

Dora, M., Kumar, M., Van Goubergen, D., Molnar, A., & Gellynck, X. (2013). Operational performance and critical success factors of lean manufacturing in European food processing SMEs. *Trends in food science & technology*, 31(2): 156-164.

Fadnavis, S., Najarzadeh, A. & Badurdeen, F. (2020). An Assessment of Organizational Culture Traits Impacting Problem Solving for Lean Transformation. *Procedia Manufacturing,* 48: 31-42.

Fagerberg, J., Fosaas, M. & Sapprasert, K. (2012). Innovation: exploring the knowledge base. *Research Policy,* 41(7): 1132-1153.

Foss, M. L., Stubbs, J. R. & Jones, G. (2011). Integrating Quality, Education, Lean, and Performance Management into a culture of continuous improvement. *Transfusion,* 51(7): 1598-1603.

Fournier, P. L., Chênevert, D. & Jobin, M. H. (2021). The antecedents of physicians' behavioral support for lean in healthcare: The mediating role of commitment to organizational change. *International Journal of Production Economics,* 232: 107961.

Fournier, P. L., Jobin, M. H., Lapointe, L., & Bahl, L. (2021). Lean implementation in healthcare: offsetting Physicians' resistance to change. *Production Planning & Control*, 1-13.

Fricke, C. F., & Buehlmann, U. (2012). Lean and Virginia's wood industry – Part II: Results and need for support. *BioResource,* 7(4), 5094-5108.

García, J. L., Rivera, L., Blanco, J., Jiménez, E., & Martínez, E. (2014). Structural equations modelling for relational analysis of JIT performance in maquiladora sector. *International Journal of Production Research*, *52*(17): 4931-4949.

Garcia-Sabater, J. & Marin-Garcia, J. A. (2011). Can we still talk about continuous improvement? Rethinking enablers and inhibitors for successful implementation. *International Journal of Technology Management,* 55: 28-42.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science,* 178(4060): 471-479.

Gholizadeh, L., Asl, I. M., Hajinabi, K. & Dehkordi, P. R. (2018). Critical success factors of lean management: An investigation of factors affecting lean management in public hospitals in Kohgiluyeh and Boyerahmad and Bushehr provinces. *International Archives of Health Sciences,* 5(4): 126-130.

Gu, Y. (2004). Global knowledge management research: a bibliometric analysis. *Scientometrics,* 61(2): 171-90.

Hama Kareem, J. A., Mohamad Al Askari, P. S. & Muhammad, F. H. (2017). Critical issues in lean manufacturing programs: A case study in Kurdish iron & steel factories. *Cogent Engineering,* 4(1): 1-1.

Hamid, R. A. & Ismail, I. R. (2016). Faktor Kejayaan Pemikiran Lean dan Pencapaian Operasi di Sektor Perkhidmatan (Success Factors of Lean Thinking and Operational Performance in Service Sector). *Jurnal Pengurusan,* 47: 1-20.

Harris, G., Stone, K. B., Mayeshiba, T., Componation, P. J., & Farrington, P. A. (2014). Transitioning from teaching lean tools to teaching lean transformation. *Journal of Enterprise Transformation*, *4*(3): 191-204.

Hayes, R., Pisano, G., Upton, D. & Wheelwright, S. (2005). Pursuing the Competitive Edge. *Wiley, New York,* pp. 1-68.

He, T., Zhang, J. & Teng, L. (2005). Basic research in biochemistry and molecular biology in China: A bibliometric analysis. *Scientometrics,* 62: 249–259.

Higgins, J. & Green, S. eds. (2004). Cochrane Reviewers' Handbook. In: *Cochrane Reviewers' Handbook 4.2.2 [updated March 2004].* Version 4.2.1 ed.: In: The Cochrane Library, Issue 1, 2004. Chichester, UK: John Wiley & Sons, Ltd.

Im, J. H. & Lee, S. M. (1989). Implementation of Just-in-time System in U.S. Manufacturing Firms. *International Journal of Operations & Production Management,* 9(1): 5–14.

Islam, A. S. M. T. (2020). Lean Fails a Lot, Even Today - Are Organizations Taking Care of All Success Factors to Implement Lean? *IUP Journal of Operations Management,* 19(2): 29-50.

Jadhav, J., Mantha, S. & Rane, S. (2014). Exploring barriers in lean implementation. *International Journal of Lean Six Sigma,* 5(2): 122-148.

Jayaraman, K., Kee, T. L. & Soh, K.L. (2012). The perceptions & perspectives of Lean Six Sigma (LSS) practitioners: An empirical study in Malaysia. *TQM Journal,* 24(5): 433-446.

Kamble, S., Gunasekaran, A. & Dhone, N.C. (2020). Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies. *International Journal of Production Research,* 58(5): 1319-1337.

Kanning, U. P. & Bergmann, N. (2009). Predictors of customer satisfaction: Testing the classical paradigms. *Managing Service Quality: An International Journal,* 19(4): 377–390.

Krishnan, C. S. N., Ganesh, L. S., & Rajendran, C. (2022). Entrepreneurial Interventions for crisis management: Lessons from the Covid-19 Pandemic's impact on entrepreneurial ventures. *International Journal of Disaster Risk Reduction*, *72*: 102830.

Kitchenham, B. & Charters, S. M. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering,* EBSE Technical Report.

Knapic, V., Jelenc, L. & Lerner, S. (2020). Hoshin Kanri - Western Management Insights on Content and Process. *FEB Zagreb International Odyssey Conference on Economics & Business,* 11: 562-578.

Knol, W. H. & Slomp, J. (2018). Implementing lean practices in manufacturing SMEs: testing ‘critical success factors’ using Necessary Condition Analysis. *International Journal of Production Research,* 56(11): 3955-3973.

Kumar, M. (2013). Adoption of Lean Manufacturing in Indian Automotive Industry: An Analysis. *IUP Journal of Mechanical Engineering,* 6(2): 15-42.

Kumar, M. (2014). Status of Quality Management practices in manufacturing SMEs: a comparative study between Australia and the UK. *International Journal of Production Research.,* 52(21): 6482-6495.

Kundu, G. & Manohar, B.M. (2012). Critical success factors for implementing lean practices in its support services. *International Journal for Quality Research,* 6(4): 301-312.

Kurpjuweit, S., Reinerth, D., Schmidt, C. G. & Wagner, S. M. (2019). Implementing visual management for continuous improvement: barriers, success factors and best practices. *International Journal of Production Research.,* 57(17): 5574-5588.

Kusnadi, A. & Yudoko, G. (2016). Contractor Work Preparation Process Improvement Using Lean Six Sigma. *South East Asian Journal of Management,* 10(1): 1-29.

Landstrom, H., Harirchi, G. & Astrom, F. (2012). Entrepreneurship: exploring the knowledge base. *Research Policy,* 41(7): 1154-1181.

Lara, G., Brasil de Brito Mello, L. C. & de Mello Bandeira, R. A. (2017). Implementation of lean manufacturing in multinational companies with production subsidiaries in Brazil: a roadmap. *European Journal of Management Studies,* 22(2): 99-123.

Laureani, A. & Antony, J. (2017). Leadership and Lean Six Sigma: a systematic literature review. *Total Quality Management and Business Excellence,* 30(1/2): 53-81.

Laureani, A. & Antony, J. (2018). Leadership – a critical success factor for the effective implementation of Lean Six Sigma. *Total Quality Management and Business Excellence,* 29(5/6): 502-523.

Li, S., Rao, S. S., Ragu-Nathan, T. S. & Ragu-Nathan, B. (2005). Development and validation of a measurement instrument for studying supply chain management practices. *Journal of Operations Management,* 23: 618–641.

Liker, J. K. (1998). Introduction: Bringing Lean back to the USA. In: J. K. Liker, ed. *Becoming Lean: Inside Stories of U.S. Manufacturers.* Portland, OR: Productivity Press: 3-40.

Liker, J. (2004). *The Toyota way: 14 management principles from the world's greatest manufacturer.* New York: McGraw-Hill.

Liker, J. K., & Convis, G. L. (2012). In J. Gallos (Ed.), Business Leadership: A Jossey-Bass Reader. New York: McGraw-Hill.

Liker, J. & Rother, M. (2011). *Why lean programs fail,* Lean Enterprise Institute: 45-79.

Lizarelli, F. L. & Alliprandini, D. H. (2020). Comparative analysis of Lean and Six Sigma improvement projects: performance, changes, investment, time and complexity. *Total Quality Management and Business Excellence,* 31(3/4): 407-428.

Lodgaard, E., Ingvaldsen, J. A., Gamme, I. & Aschehoug, S. (2016). Barriers to Lean Implementation: Perceptions of Top Managers, Middle Managers and Workers. *Procedia CIRP,* 57: 595-600.

Madhani, P. M. (2020). Performance Optimisation of Retail Industry: Lean Six Sigma Approach. *ASBM Journal of Management,* 13(1/2): 74-91.

Mann, D. (2009). The missing link: Lean leadership. *Frontiers of Health Services Management,* 26(1): 15 –26.

Manville, G., Greatbanks, R., Krishnasamy, R. & Parker, D. W. (2012). Critical success factors for Lean Six Sigma programmes: a view from middle management. *International Journal of Quality and Reliability Management,* 29(1): 7-20.

Marin-Garcia, J. A. & Bonavia, T. (2015). Relationship between employee involvement and lean manufacturing and its effect on performance in a rigid continuous process industry. *International Journal of Production Research,* 53(11): 3260-3275.

Markos, S., & Sridevi, M. S. (2010). Employee engagement: The key to improving performance. *International Journal of Business and Management*, *5*(12): 89.

Marodin, G. A. & Saurin, T. A. (2012). Implementing lean production systems: research areas and opportunities for future studies. *International Journal of Production Research,* 51(22): 6663-6680.

Marodin, A. G. & Saurin, T. A. (2015). Managing barriers to lean production implementation: context matters. *International Journal of Production Research,* 53(13): 3947-3962.

Marodin, G. A. & Saurin, T. A. (2015). Classification and relationships between risks that affect lean production implementation: A study in southern Brazil. *Journal of Manufacturing Technology Management,* 26(1): 57-79.

Martinez-Jurado, P.J. & Moyano-Fuentes, J. (2012). Key determinants of lean production adoption: evidence from the aerospace sector, *Production Planning & Control*, 25(4): 332-345.

Meredith, J. R.& McTavish, R. (1992). Organized manufacturing for superior market performance. *Long Range Planning,* 25(6): 63–71.

Minh, N., Tuấn, L. & Nguyen, D. N. (2015). Framework of critical success factors for lean implementation in Vietnam manufacturing enterprise. *VNU Journal of Science: Economics and Business,* 31(5E): 33-41.

Mirdad, W. K. & Eseonu, C. I. (2015). A conceptual map of the lean nomenclature: comparing expert classification to the lean literature. *Engineering Management Journal,* 27(4): 188-202.

Moacir, G. F. (2016). Lean manufacturing in Brazilian small and medium enterprises: implementation and effect on performance. *International Journal of Production Research,* 54(24): 7523-7545.

Moed, H. F. (2005). *Citation analysis in research evaluation.* Springer.

Moeuf, A., S. Tamayo, S. Lamouri, R. Pellerin, & A. Lelievre. (2016). Strengths and weaknesses of small and medium sized enterprises regarding the implementation of lean manufacturing. *IFAC-PapersOnLine* 49: 71–76.

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., & Barbaray, R. (2018). The industrial management of SMEs in the era of Industry 4.0. *International Journal of Production Research,* 56(3): 1118-1136.

Mohanty, R. P., Yadav, O. P. & Jain, R. (2007). Implementation of lean manufacturing principles in auto industry. *Vilakshan-XIMB Journal of Management,* 1-32.

Monden, Y. (1998). *Toyota Production System: An Integrated Approach to Just-in-Time.* Norcross, GA: Engineering & Management Press.

Narasimhan, R., Talluri, S. & Mahapatra, S. K. (2006). Multiproduct, multicriteria model for supplier selection with product life-cycle considerations. *Decision Sciences,* 37(4): 577–603.

Narkhede, B. E., Raut, R. D., Roy, M., Yadav, V. S., & Gardas, B. (2020). Implementation barriers to lean-agile manufacturing systems for original equipment manufacturers: an integrated decision-making approach. *International Journal of Advanced Manufacturing Technology*, 108(9): 3193-3206.

Netland, T. (2016). Critical success factors for implementing lean production: the effect of contingencies. *International Journal of Production Research,* 54 (8): 2433-2448.

Niewiadomski, P., Pavlak, N. & Tsimayeu, A. (2018). Barriers to effective implementation of lean management principles - empirical exemplification in the industry of agricultural machinery. *Log Forum,* 14(4): 563-576.

Noori, B. (2015). Identifying Critical Issues in Lean Implementation in Hospitals. *Hospital Topics,* 93(2): 44-52.

Olivella, J., Cuatrecasas, L. & Gavilan, N. (2007). Work organisation practices for lean production. *Journal of Manufacturing Technology Management,* 19(7): 798-811.

Pakdil, F. & Leonard, K. M. (2014). Criteria for a lean organisation: development of a lean assessment tool. *International Journal of Production Research,* 52(15): 4587-4607.

Pakdil, F. & Leonard, K. M. (2017). Implementing and sustaining lean processes: the dilemma of societal culture effects. *International Journal of Production Research,* 55(3): 700-717.

Panizzolo, R. (1998). Applying the lessons learned from 27 lean manufacturers: the relevance of relationships management. *International Journal Production Economics,* 55: 223–240.

Papadopoulou, T. C. & Özbayrak, M. (2005). Leanness: experiences from the journey to date. *Journal of Manufacturing Technology Management,* 16(7): 784-807.

Paranitharan, K. P., Ramesh, B. T., Pal, P. A. & Jeyathilagar, D. (2017). An empirical validation of integrated manufacturing business excellence model. *International Journal of Advanced Manufacturing Technology,* 92(5-8): 2569-2591.

Pattnaik, D., Hassan, M. K., Kumar, S., & Paul, J. (2020). Trade credit research before and after the global financial crisis of 2008–A bibliometric overview. *Research in International Business and Finance*, 54: 101287.

Pedro, G.-V., Macarena, S.-D. & Luis Miguel, M.-R. (2016). Six Sigma in SMES with low production volumes. A successful experience in aeronautics. *Universia Business Review,* 51: 52-71.

Pereira, L. & Tortorella, G. (2018). Identification of the Relationships Between Critical Success Factors, Barriers and Practices for Lean Implementation in a Small Company. *Brazilian Journal of Operations and Production Management,* 15(2): 232-246.

Pingyu, Y., & Yu, Y. (2010). The barriers to SMEs' implementation of lean production and countermeasures: based on SMS in Wenzhou. *International Journal of Innovation, Management and Technology*, 1(2): 220-225.

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P. & Bachrach, D. G. (2008). Scholarly influence in the field of management: a bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management,* 34(4): 641-720.

Poksinska, B., Swartling, D. & Drotz, E. (2013). The daily work of Lean leaders – Lessons from manufacturing and healthcare. *Total Quality Management & Business Excellence,* 24(7-8): 886 –898.

Popoola, A. A. (2000). Development of a methodology for the rapid implementation of a sustainable lean manufacturing system. Thesis, Massachusetts Institute of Technology, Sloan School of Management, Cambridge, MA.

Prashar, A. (2018). Toward cycle time reduction in manufacturing SMEs: Proposal and evaluation. *Quality Engineering,* 30(3): 469-484.

Pritchard, A. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation,* 25: 348-349.

Pulsiri, N., & Vatananan-Thesenvitz, R. (2018). Improving systematic literature review with automation and bibliometrics. In *2018 Portland International Conference on Management of Engineering and Technology (PICMET)* (pp. 1-8). IEEE.

Raja, S., Raju, R., Rajkanth, R. & Nagaraj, M. (2018). An empirical assessment of Lean Six Sigma Awareness in manufacturing industries: construct development and validation. *Total Quality Management and Business Excellence,* 29(5/6): 686-703.

Repenning, N. P. & Sterman, J. D. (2001). Nobody ever gets credit for fixing problems that never happened: Creating and sustaining process improvement. *California Management Review,* 43(4): 64–88.

Rockart, J.F. (1979). Chief executives define their own data needs. *Harvard Business Review*, 57(2): 81-93.

Rose, A. N. M., Deros, B. & Ab. Rahman, M. N. (2013). A study on lean manufacturing implementation in Malaysian automotive component industry. *International Journal of Automotive and Mechanical Engineering,* 8: 1467-1476.

Saks, A. M. (2006). Antecedents and consequences of employee engagement. *Journal of Managerial Psychology,* 21(6): 600-619.

Salem, R., Musharavati, F., Hamoud, A. M. & Al-Khalifa, K. N. (2016). An empirical study on lean awareness and potential for lean implementations in Qatar industries. *The International Journal of Advanced Manufacturing Technology,* 82: 1607–1625.

Salma, A., Anas, C., & Mohammed, E. H. (2021). Bibliographic study on the difficulties encountered by SMEs during the implementation of lean manufacturing. *Journal of Advanced Manufacturing Systems*, 20(01): 163-190.

Sanjay, B. (2012). Prominent obstacles to lean. *International Journal of Productivity and Performance Management,* 61(4): 403-425.

Scalera, F. (2011). World class manufacturing and "Six Sigma": Their application to SMEs, as well as to services. *Manager (University of Bucharest, Faculty of Business & Administration),* 14: 57-75.

Secchi, R. & Camuffo, A. (2016). Rolling out lean production systems: a knowledge-based perspective. *International Journal of Operations & Production Management*, 36(1): 61-85.

Seskir, Z. C. & Aydinoglu, A. U. (2021). The landscape of academic literature in quantum technologies. *International Journal of Quantum Information,*19(2): 2150012.

Shah, R. & Ward, P. T. (2003). Lean manufacturing: context, practice bundles, and performance. *Journal of Operations Management,* 21(2): 129-149.

Shah, R. & Ward, P. T. (2007). Defining and developing measures of lean production. *Journal of Operations Management,* 25: 785–805.

Shamsudin, S., Mohd Radzi, N. I. & Othman, R. (2016). Causal ambiguity in lean production implementation in Malaysia. *Journal of Asia-Pacific Business,* 17(3): 249-266.

Sim, K. L. & Chiang, B. (2012). Lean production systems: resistance, success and plateauing. *Review of Business,* 33(1): 97-110.

Singh, B., Garg, S. K. & Sharma, S. K. (2010). Scope for lean implementation: a survey of 127 Indian industries. *International Journal of Rapid Manufacturing,* 1(3): 323-333.

Sisson, J. & Elshennawy, A. (2015). Achieving success with Lean: An analysis of key factors in Lean transformation at Toyota and beyond. *International Journal of Lean Six Sigma,* 6(3): 263-280.

Skrudupaite, A. & Jucevicius, R. (2011). Critical success factors for the implementation of the synchronized production system. *Social Sciences,* 71(2): 16-23.

Sohal, A. & Eggleston, A. (1994). Lean production: experience amongst Australian organisations. *International Journal of Operations & Production Management,* 14: 1-17.

Soni, G. & Kodali, R. (2016). Interpretive structural modeling and path analysis for proposed framework of lean supply chain in Indian manufacturing industry. *Journal of Industrial and Production Engineering,* 33(8): 501-515.

Sparrow, P. & Otaye-Ebede, L. (2014). Lean management and HR function capability: the role of HR architecture and the location of intellectual capital. *International Journal of Human Resource Management,* 25(21): 2892-2910.

Srimathi, K. & Narashiman, K. (2021). Leadership Styles and Their Impact on Lean Six Sigma Practices in Indian Industries. *South African Journal of Industrial Engineering,* 32(1): 1-12.

Stankalla, R., Koval, O. & Chromjakova, F. (2018). A review of critical success factors for the successful implementation of Lean Six Sigma and Six Sigma in manufacturing small and medium sized enterprises. *Quality Engineering,* 30(3): 453-468.

Taherimashhadi, M. & Ribas, I. (2018). A Model to align organizational culture to lean culture. *Journal of Industrial Engineering and Management,* 11(2): 207-221.

Taylor, A., Taylor, M. & McSweeney, A. (2013). Towards greater understanding of success and survival of lean systems. *International Journal of Production Research,* 51(22): 6607-6630.

Thirkell, E. & Ashman, I. (2014). Lean towards learning: connecting Lean Thinking and human resource management in UK higher education. *International Journal of Human Resource Management,* 25(21): 2957-2977.

Timans, W., Antony, J., Ahaus, K., & van Solingen, R. (2012). Implementation of Lean Six Sigma in small-and medium-sized manufacturing enterprises in the Netherlands. *Journal of the Operational Research Society*, *63*(3): 339-353.

Tortorella, G. L., Marodin, G. A., Fogliatto, F. S. & Miorando, R. (2015). Learning organisation and human resources management practices: an exploratory research in medium-sized enterprises undergoing a lean implementation. *International Journal of Production Research,* 53(13): 3989-4000.t

Tran, D. T., Pham, H. T., & Bui, V. T. (2020). The effect of contextual factors on resistance to change in lean transformation. *The Journal of Asian Finance, Economics and Business*, *7*(11): 479-489.

Turesky, E. F. & Connell, P. (2010). Off the rails: understanding the derailment of a lean manufacturing initiative. *Organization Management Journal,* 7: 110-132.

Ulewicz, R. & Kuceba, R. (2016). Identification of problems of implementation of Lean concept in the SME sector. *Economics & Management,* 8(1): 19-25.

Vallejo, V. F., Antony, J., Douglas, J. A., Alexander, P., & Sony, M. (2020). Development of a roadmap for Lean Six Sigma implementation and sustainability in a Scottish packing company. *The TQM Journal*, *32*(6): 1263-1284.

van der Merwe, K. R., Pieterse, J. J. & Lourens, A. S. (2014). The development of a theoretical lean culture causal framework to support the effective implementation of lean in automotive component manufacturers. *South African Journal of Industrial Engineering,* 25(1): 131-144.

van Eck, N. J. & Waltman, L. (2013). *VOSviewer.* [Online]
Available at: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.5.4.pdf
[Accessed December 2021].

Wang, Y., & Mi, J. (2012). Searchability and discoverability of library resources: Federated search and beyond. *College & Undergraduate Libraries*, *19*(2-4): 229-245.

White, R. E., Pearson, J. N. & Wilson, J. R. (1999). JIT Manufacturing: A survey of implementation in small and large US manufacturers. *Management Science,* 45: 1–15.

Wilson, L. (2010). *How to Implement Lean Manufacturing.* New York: Mc-Graw Hill.

Womack, J. P. & Jones, D. T. (1996). *Lean Thinking.* New York, NY: Simon and Schuster.

Womack, J. P., Jones, D. T. & Roos, D. (1990). *The Machine that Changed the World.* New York: Rawson Associates.

Yuik, C. J. & Puvanasvaran, P. (2020). Development of lean manufacturing implementation framework in machinery and equipment SMEs. *International Journal of Industrial Engineering and Management,* 11(3): 157-169.

Yuik, C. J., Perumal, P. A., & Feng, C. J. (2020). Exploring critical success factors for the implementation of lean manufacturing in machinery and equipment SMEs. *Engineering Management in Production and Services*, *12*(4): 77-91.

Zargun, S. & Al-Ashhab, A. (2014). Critical success factors for lean manufacturing: a systematic literature review - an international comparison between developing and developed countries. *Advanced Materials Research,* 845: 668-681.

Zhang, B., Niu, Z., Feng, L. & Wang, X. (2020). Evaluation system for lean knowledge management ability based on improved grey correlation analysis. *International Journal of Industrial Engineering,* 27(5): 712-730.

Zhang, B., Niu, Z., & Liu, C. (2020). Lean tools, knowledge management, and lean sustainability: the moderating effects of study conventions. *Sustainability*, *12*(3): 956.

.

Appendix 1: The first four CSFs

CSFs	The terms	(Frequencies)
Management Commitment	Management engagement and commitment, the attention and involvement of senior managers, clear motivation to implement, commitment of top management (Antony, et al., 2008); (Hamid and Ismail, 2016); (Prashar, 2018); (Manville, et al., 2012); (Brun, 2011); (Scalera, 2011); (Antony and Desai, 2009); (Kurpjuweit, et al., 2019); (Lara, et al., 2017); (Arabi, et al., 2021); (Stankalla, et al., 2018); (Netland, 2016); (Baskaran and Lakshmanan, 2019); (Pereira and Tortorella, 2018); (Raja, et al., 2018); (Knol and Slomp, 2018); (Hama Kareem, et al., 2017); (Berlec, et al., 2017); (Belhadi, et al., 2016); (Badgujar, et al., 2016); (Jayaraman, et al., 2012); (Pingyu and Yu, 2010); (Turesky and Connell, 2010) (Antony and Kumar, 2012); (Garcia-Sabater and Marin-Garcia, 2011); (Papadopoulou and Özbayrak, 2005); (Bagni, et al., 2021); (Paranitharan, et al., 2017); (Skrudupaite and Jucevicius, 2011); (Belhadi et al., 2019) (Yuik et al., 2020) (Timans et al., 2012)	(32, 8.7%)
Leadership	Leadership, contingent on the situation (contingency theory), the ownership of a factory by a specific corporation, financial accountability, reduce leadership conflict, (participative, supportive, instrumental) leadership, leadership and management, accountability, ownership, follow-up, leader-member exchange, lean leadership, (leadership for SS) (Turesky and Connell, 2010); (Srimathi and Narashiman, 2021) (Shamsudin, et al., 2016) (Dombrowski and Mielke, 2013) (Antony and Desai, 2009) (Yuik and Puvanasvaran, 2020) (Gholizadeh, et al., 2018) (Lara, et al., 2017) (Arabi, et al., 2021) (Netland, 2016) (Achanga, et al., 2006) (Islam, 2020) (Pereira and Tortorella, 2018) (Laureani and Antony, 2017) (Knol and Slomp, 2018) (Taylor, et al., 2013) (Antony and Kumar, 2012) (Garcia-Sabater and Marin-Garcia, 2011) (Bagni, et al., 2021) (Dehdasht, et al., 2020) (Paranitharan, et al., 2017) (Kumar, 2014) (Kusnadi and Yudoko, 2016) (Yuik et al., 2020)	(24, 6.5%)
Employee commitment and engagement	Desire to improve service, team's commitment, employee engagement and involvement, improve the commitment of employees, motivate employees and shape their behaviour, commitment, and motivation of employees (Turesky and Connell, 2010) (Pedro, et al., 2016) (Hamid and Ismail, 2016) (Deflorin and Scherrer-Rathje, 2012) (Yuik and Puvanasvaran, 2020) (Tortorella, et al., 2015) (Marin-Garcia and Bonavia, 2015) (Dombrowski and Mielke, 2013) (Yuik and Puvanasvaran, 2020) (Kurpjuweit, et al., 2019) (Pereira and Tortorella, 2018) (Pakdil and Leonard, 2017) (Alhuraish, et al., 2017) (Belhadi, et al., 2016) (Pingyu and Yu, 2010) (Antony and Kumar, 2012) (Garcia-Sabater and Marin-Garcia, 2011) (Olivella, et al., 2007) (Bagni, et al., 2021) (Dehdasht, et al., 2020) (Skrudupaite and Jucevicius, 2011) (Yuik et al., 2020)	(22, 6.0%)

Knowledge of lean tools	Understanding of SS methodology, learning lean principles and tools, prior learning, guidelines deployment, proper methodology of implementation, value stream mapping, visual management, facility of understanding the concepts of lean construction (Antony, et al., 2008) (Moacir, 2016) (Pedro, et al., 2016) (Tortorella, et al., 2015) (Manville, et al., 2012) (Brun, 2011) (Lizarelli and Alliprandini, 2020) (Arabi, et al., 2021) (Stankalla, et al., 2018) (Netland, 2016) (Pereira and Tortorella, 2018) (Belhadi, et al., 2016) (Sisson and Elshennawy, 2015) (Jayaraman, et al., 2012) (Garcia-Sabater and Marin-Garcia, 2011) (Olivella, et al., 2007) (Papadopoulou and Özbayrak, 2005) (Bagni, et al., 2021) (Dehdasht, et al., 2020) (Skrudupaite and Jucevicius, 2011), (Bhadu et al., 2021)	(21, 5.7%)

Appendix 2: The first five barriers

Barriers	The terms	(Frequencies)
Resistance to change	Internal resistance, resistance to change from employees, appetite for change, employee attitudes, resistance to change, operators are insecure in carrying out new attributions, and resistance to change offered by the workers and supervisors (Ayçin, 2021) (Abolhassani and Gopalakrishnan, 2016) (Kurpjuweit, et al., 2019) (Salem, et al., 2016) (Soni and Kodali, 2016) (Arabi, et al., 2021) (Narkhede, et al., 2020) (Pereira and Tortorella, 2018) (Marodin and Saurin, 2015) (Taylor, et al., 2013) (Bhasin, 2012) (Antony and Kumar, 2012) (Antony, et al., 2008) (Salma et al., 2021) (Timans et al., 2012)	(15, 10.5%)
Lack of training and knowledge	Chosen lean tools and practices not according to the best practice, preferred lean tools and practices not adding sufficient value, failure to prioritise lean tools and practices, lack of technical knowledge about lean methodologies, ignorance of lean tools, lack of OEE knowledge, lack of clarity related to lean concepts, lack of knowledge about lean, lack of training and knowledge about lean tools, lack of knowledge about SS methodologies (Ayçin, 2021) (Antony, et al., 2008) (Abolhassani and Gopalakrishnan, 2016) (Ulewicz and Kuceba, 2016) (Lara, et al., 2017) (Soni and Kodali, 2016) (Arabi, et al., 2021) (Lodgaard, et al., 2016) (Mirdad and Eseonu, 2015) (Marodin and Saurin, 2012) (Antony and Kumar, 2012) (Salma et al., 2021)	(12, 8.4%)
Lack of leadership commitment and support	Lack of leadership commitment and support, lack of commitment of top and middle management support (Ayçin, 2021) (Bevilacqua, et al., 2017) (Lara, et al., 2017) (Salem, et al., 2016) (Soni and Kodali, 2016) (Arabi, et al., 2021) (Lodgaard, et al., 2016) (Marodin and Saurin, 2015) (Ulewicz and Kuceba, 2016) (Marodin and Saurin, 2015) (Salma et al., 2021) (Timans et al., 2012)	(12, 8.4%)
Lack of resources	Lack of resources (financial and human etc.), resource constraints, insufficient management time, investment restrictions, lack of human and/or financial resources (Ayçin, 2021) (Antony, et al., 2008) (Bevilacqua, et al., 2017) (Kurpjuweit, et al., 2019) (Soni and Kodali, 2016) (Baskaran and Lakshmanan, 2019) (Bhasin, 2012) (Antony and Kumar, 2012) (Timans et al., 2012)	(9, 6.3%)
Lack of financial resources	Initial capital, budgetary, financial constraints, lack of finance, cost of investment, insufficient external and internal funding, and investment restrictions (Salem, et al., 2016) (Soni and Kodali, 2016) (Arabi, et al., 2021) (Narkhede, et al., 2020) (Baskaran and Lakshmanan, 2019) (Marodin and Saurin, 2015) (Bhasin, 2012) (Salma et al., 2021)	(8, 5.6%)

DECISION SCIENCES INSTITUTE

The Effects of Personalized Push Notifications on Mobile Shoppers' Attitude Towards Push Notification Engagement and Acceptance.

Mbuso Levon Gama
National Cheng Kung University
Email: mbusogama@gmail.com

Jengchung Victor Chen
National Cheng Kung University
Email: victor@mail.ncku.edu.tw

ABSTRACT

Online companies usually deploy push notifications to disseminate marketing information, negating the anxiety about information arrival. Concurrently, push notification increases constant checking behavior and suppress personal productivity. Furthermore, the intrusive nature of push notifications amplifies frustration and leads to passive push notification acceptance. Our study proposes personalizing push notifications to negate push notifications frustration and increase attitudes toward push notifications. Using a 280-sample size, we investigate the effect of push notifications' personalization. Structural equation model results indicate support for our assumption. Specifically, charity appeals, scarcity messages, and price incentives are associated with need-based and location-based matching, increasing attitudes toward push notifications. Findings and recommendations are discussed.

KEYWORDS: Constant checking, Passive acceptance, Personalization, Intrusiveness, and Attitudes

INTRODUCTION

In the contemporary era, the widespread utilization of smartphones has led to an abundance of accessible information, accessible anytime and anywhere worldwide through mobile applications (Hussain et al., 2017). This extraordinary accessibility to smartphones has cultivated a profound attachment of mobile phone users to their devices. This phenomenon is referred to as technological dependency, instilling a fear of smartphone withdrawal(Tams et al., 2018). Moreover, the extensive usage of smartphones has entrapped users in a state of technological entrenchment, where they find themselves unable to regulate or control their smartphone usage, even when it contradicts their productivity objectives. Numerous studies indicate that in the United States, individuals check their smartphones approximately every 12 minutes and engage with them around 80 times per day (Iyer & Zhong, 2022). Additionally, previous research reveals a consistent surge in smartphone use among Americans, projecting that the average American smartphone user spends at least 3.3 hours daily interacting with their device (Pielot et al., 2014). Evidently, this signals a concerning pattern of excessive smartphone use among individuals, given the ongoing demands of daily tasks and responsibilities.

Consequently, the prevalent reliance on smartphones grants users access to copious amounts of information at their fingertips, enabling online companies to utilize push notifications as a means to deliver information to users promptly. While push notifications offer users the advantage of receiving real-time updates, alleviating the anxiety associated with waiting for information (Iyer & Zhong, 2022), they also contribute to the development of constant checking behavior. This behavior poses a challenge to users as it disproportionately increases their time spent online. The inclination to check for updates frequently stems from an inability to tolerate uncertainty (Tolin et al., 2003), resulting in reduced personal productivity due to time lost to this behavior (Duke & Montag, 2017).

Push notifications serve as a valuable tool for developers to deliver essential information to users at any time, thereby enhancing the overall value of their apps(Gavilan & Martinez-Navarro, 2022; Liu et al., 2022; März et al., 2021). However, the improper implementation of push notifications can lead to issues. Push notifications issues range from the potential risk of mismanagement, such challenges encompass issues like poorly timed message delivery, a perceived lack of value in the content, inappropriate message content, and an overwhelming frequency of messaging. When these challenges arise, push notifications risk becoming intrusive, bothersome, and unwelcome, which in turn can have a negative impact on user experience and ultimately diminish users' attachment to the brand(Gavilan & Martinez-Navarro, 2022). Excessive notifications have the potential to prompt users to uninstall applications, consequently elevating the mobile app user churn rate. On average, the typical US smartphone user receives around 46 push notifications daily. Reports indicate that 10% of users deactivate push notifications in response to a weekly notification, while 40% opt out when faced with 3-5 messages per week. Interestingly, only 5% of users turn off push notifications after encountering over 20 push messages (Dogtiev, 2023; HelpLama, 2023). The reception of push notifications hinges significantly on message content; users tend to be more tolerant of notifications that align with their interests. Additional reports highlight user perspectives on bothersome push notifications: 53% find them irritating, 49% believe they cause distractions, 39% feel notifications disrupt them at inopportune times, and 24% and 12% of users respectively deem push notifications irrelevant and impersonal(Dogtiev, 2023).

Amid the escalating frustration caused by push notifications, this study introduces the concept of personalizing push notifications through need-based and location-based matching. This approach aims to mitigate push notification frustration and enhance users' willingness to engage with and accept push notifications. Chen et al. (2019) assert that the intrusive nature of push notifications contributes to this frustration, often leading to passive acceptance of these notifications. The escalation in the frequency of non-personalized push notifications has been observed to correlate with a rise in app uninstalls while concurrently leading to a decline in the direct open rate of push notifications (Wohllebe et al., 2021). However, studies indicate that notifications strategically timed to coincide with periods of potential user disengagement can have a positive impact on user engagement. By delivering notifications to users when they are most susceptible to disengagement, there is an increase in their perception of the app's utility (Bell et al., 2023). While push notifications are not outrightly rejected, users develop coping strategies to manage their annoyance, recognizing the significance of this information in achieving their online objectives. To this end, to the best of our knowledge, no empirical study has been conducted that integrates both promotional activities and personalization techniques (need-based and location-based matching) to examine users' attitudes regarding the engagement with and acceptance of push notifications. Need-based matching involves tailoring marketing content according to an individual's personal requirements, whereas location-based matching tailors marketing information based on the appropriate context for its reception(Ho & Lim, 2018). Accordingly, we propose personalizing three types of promotions (charity appeals, scarcity messages, and price incentives) to yield more effective outcomes. Furthermore, building on previous research that underscores the influence of consumer mood on decision-making in personalized recommendations, we aim to address the following two questions:

RQ1: To what extent does the personalization of push notification services from promotional activities (charity appeals, scarcity messages, and price incentives) influence mobile shoppers' attitudes towards engaging and accepting push notifications?

RQ2: How does consumer mood moderate the relationship between promotional interventions (charity appeals, scarcity messages, and price incentives) and personalization (need-based and location-based matching) in online shopping?

Prior research has delved into the strategies that users and designers adopt to navigate push notification frustration (Chen et al., 2019), the implementation of privacy protection functions in push notifications (Cheng et al., 2022), the utilization of progression and social incentives in gamified push notifications (Kunkel et al., 2023), enhancing the quality of the information provided to users in notification (Gavilan & Avello, 2023), and the role of personalization in unplanned purchase behavior and personalization in unplanned purchases(Ho et al., 2011; Ho & Lim, 2018). In the current study, we present a novel approach to address the frustration experienced by users in response to push notifications. While existing literature suggests that push notifications alleviate information-related anxiety, they simultaneously contribute to the surge in constant checking behavior (Iyer & Zhong, 2022). Our study lays the groundwork for mitigating this phenomenon of incessant checking. In doing so, we make a significant contribution to the field of advertising literature. By proposing and examining the integration of personalized push notifications, we offer valuable insights into the potential efficacy of such tailored marketing interventions. This study introduces a progressive perspective that could reshape how push notification engagement is approached, ultimately providing evidence for the utility of personalization as a tool for effective marketing strategies.

THEORETICAL BACKGROUND

Push notifications and online shopping.

The widespread use of smartphones has led to the availability of a vast amount of information accessible anytime and anywhere around the globe through mobile applications (Hussain et al., 2017). Numerous online companies utilize push notifications to distribute marketing content to online consumers. While these marketing strategies offer benefits by alleviating the uncertainty of information delivery, they also come with certain drawbacks, such as fostering a habit of frequent checking behavior (Iyer & Zhong, 2021). This continuous checking behavior involves the constant quest for information and is akin to pathological doubt, stemming from an intolerance of uncertainty (Tolin et al., 2003). Push notifications, in their attempt to deliver substantial information intrusively (Chen et al., 2019), can exacerbate this checking behavior. Furthermore, this persistent checking habit proves detrimental to personal productivity due to the frequent disruptions users face (Duke & Montag, 2017).

In the context of online shopping Chen et al. (2019) indicate that mobile shoppers often find themselves overwhelmed by information overload and the intrusive nature of push notifications. Amidst this frustration, users develop coping mechanisms, such as managing disturbances and self-preservation techniques, to contend with the constant barrage of push notifications. Users tend to view this information as relevant to their shopping objectives rather than dismissing it outright (Chen et al., 2019). Supporting this reasoning are the findings of Tams et al. (2018), who highlight that reliance on technology gives rise to a social fear known as "nomophobia," which pertains to the anxiety surrounding smartphone separation. Nonetheless, in the long run, users are projected to take measures such as blocking bothersome push notifications (Iyer & Zhong, 2022) and adopting self-preservation strategies (Chen et al., 2019). However, blocking or discontinuing push notifications could pose challenges for app designers, practitioners, and marketers who rely on user engagement with the information presented.

Our study proposes a solution involving personalized push notifications based on location and needs matching to mitigate anxiety and perpetual information checking. Tailoring push notification messages—such as messages pertaining to scarcity, promotional offers, and charitable appeals—using location and need-based matching can alleviate the frustration associated with push notifications, as users receive them when they anticipate such content. Narang and Shankar (2019) highlight that app link services enable shoppers to access apps when they are near a store, providing them with loyalty rewards, notifications, and product details. This approach helps mitigate concerns related to the fear of missing out (FOMO), as users become accustomed to receiving information aligned with their location and individual needs

Consumer shopping goal stages theory

According to Lee and Ariely (2006), the consumer shopping process can be delineated into two distinct stages. Drawing from the concept of goals, they proposed a two-stage framework. Consumers generally hold abstract goals in the initial stage, which are more receptive to external cues due to the prevailing uncertainty surrounding their goals. During this initial phase, consumers gather information and explore product options aligned with their preferences. At this juncture, consumers often lack clarity regarding the number of products they desire and their intended expenditure. Consequently, they are susceptible to various types of information during the early stages of their purchasing journey. In the subsequent stage, goals tend to be specific and less influenced by external cues. Consumers typically possess well-defined and concrete goals during this phase, necessitating different considerations. By this point, consumers have greater certainty about their preferred products and have established a clear budget for spending. Thus, consumers are more likely to reach a purchasing decision in this later stage, and marketing interventions are deemed less impactful at this juncture. In essence, the implications for marketing are contingent on the consumer's position within their purchasing journey.

Numerous studies corroborate the existence of distinct stages within the consumer shopping process. For instance, Song et al. (2017) found that recommendation messages from weak ties are more effective in the initial stage, whereas messages from strong ties prove more influential in the later stages of the shopping process, particularly when deal scarcity is low. Similarly, Tam and Ho (2005) assert that the persuasion process transpires at different points during a consumer's shopping journey, as supported by their empirical investigation into web personalization. Interrupting-based advertising, such as pop-ups, is more effective when designed with implicit advertising intent in the pre-decisional stage, as demonstrated by Chan et al. (2009). than when explicit advertising intent is employed. Additionally, promotional strategies vary according to the type of product. Bridges and Florsheim (2008) discovered that incorporating utilitarian flow elements in advertising enhances purchase intention, whereas no discernible correlation exists between hedonic flow elements and online purchases. Contemporary studies further align with prior research, illustrating the effectiveness of promotional activities that align with shopping goals across different stages of the shopping process (Berger & Burkhalter, 2022; Trzebinski et al., 2023). This study reveals a positive relationship between a goal-oriented mindset and preference descriptions centered on benefits, as opposed to attribute-based descriptions. The body of research consistently supports the proposed consumer shopping goal stages framework.

Push notification personalization.

To tailor push notifications effectively, a thorough understanding of users' context is paramount in order to align with their mental representations (Ho & Lim, 2018). This necessitates a grasp of context awareness achieved through meticulous curation of user preferences. Preferences can be defined as outcomes arising from the acquisition and upkeep of a particular object or event (Zajonc & Markus, 1982). Importantly, preferences are not inherent but rather are developed through elicitation. The introduction and exposure to various alternatives or choices guide users in shaping their preferences. For example, a user might develop a preference for a MacBook laptop over other brands after having the opportunity to use it. This exposure aids users in constructing well-defined preferences. Consequently, promotional activities play a significant role in facilitating the development of such preferences. , studies indicate that notifications strategically timed to coincide with periods of potential user disengagement can have a positive impact on user engagement. By delivering notifications to users when they are most susceptible to disengagement, there is an increase in their perception of the app's utility (Bell et al., 2023). This highlights the significance of timing in push notifications, suggesting that well-time notifications can serve as effective prompts for user re-engagement**.**In the context of mobile push notifications, their effectiveness in promoting engagement is evident. Push notifications have been found to stimulate redemption behavior among program participants, thereby enhancing various metrics such as total stamp collection and spending within a specific chain or context (Bies et al., 2021).

This underscores the potential of push messaging to drive not only engagement but also tangible actions and behaviors from users. Synthesizing arguments for the use of location-based push notifications reveals several compelling dimensions. Firstly, by leveraging location data, push notifications can be tailored to encourage offline visits, effectively bridging the gap between the digital and physical realms. Secondly, the strategic deployment of coupon offers through push notifications has been shown to enhance user response rates, as they perceive tangible value in these personalized incentives. Thirdly, the delivery of personalized and high-engagement content via push notifications can foster a sense of individual relevance and exclusivity, leading to heightened engagement levels. Collectively, these dimensions underline the diverse capabilities of location-based push notifications in capturing user attention, driving behavior, and fostering a sense of connection between the user and the application(Kamiya & Branisso, 2021).

HYPOTHESES DEVELOPMENT AND RESEARCH MODEL

We base our research model on the foundational notion that personalizing promotional activities in online shopping through location-based and need-based matching can alleviate the frustration and anxiety induced by push notifications. This personalized approach is anticipated to encourage active acceptance of push notifications rather than mere passive acceptance, as highlighted by Chen et al. (2019)**.** We posit that implementing location-based and need-based matching will enhance users' willingness to engage with push notifications. Additionally, drawing from the mood congruence theory, we propose that users' emotional states or moods will moderate the connection between promotional activities (charity appeals, scarcity messages, and price incentives) and the effectiveness of location-based and need-based matching.

In our conceptualization, charity appeals refer to advertisements that emphasize the contribution of proceeds from specific product sales toward uplifting individuals or societal groups facing challenging circumstances. Charity appeals are categorized into two dimensions: Negative appeals, which emphasize the potential adverse consequences for deprived individuals or societies if product sales fall short, and positive appeals, which highlight the benefits anticipated to result from successful product sales(Choi et al., 2016; Erlandsson et al., 2018), positive appeals enhance favorable attitudes towards complying with charity appeals, while negative appeals prompt actual compliance with charity appeals.

However, negative emotional appeals lacking credible descriptive rationales are found to be ineffective(Majumdar & Bose, 2018). Charity appeals aligned with cognitive processing tend to be more influential in shaping users' decision-making (Lee et al., 2020; Septianto & Tjiptono, 2019). Charity appeals that evoke feelings of nostalgia are also more likely to impact users' decisions and intentions (Ford & Merchant, 2010; Kim & Childs, 2021). According to the consumer shopping goal stage theory proposed by Lee and Ariely (2006), marketing activities that resonate with consumers' shopping objectives tend to be more effective. Location-based advertising is favored and holds efficacy in influencing consumer purchase intentions (Molitor et al., 2020; Saputra & Setiawan). Similarly, Ho and Lim (2018) demonstrate that location-based and need-based matching personalization evokes unplanned purchases. Thus, we hypothesize:

> *H1a: Charity appeals are positively associated with need-based matching personalization of push notifications.*
> *H1b: Charity appeals are positively associated with location-based matching personalization of push notifications.*

Scarcity messages serve as a common strategy to raise awareness about either perceived or actual product shortages. In certain cases, these messages are used to notify consumers about the limited availability period of a product (Ladeira et al., 2023). Scarcity messages have proven effective in eliciting purchase intention, particularly for highly sought-after products, where encountering a scarcity message amplifies the intention to purchase a scarce product (Aggarwal et al., 2011; Fogel & Setton, 2022; Ladeira et al., 2023; Song et al., 2019). Hmurovic et al. (2023) contend that time scarcity promotions within the online environment may yield comparable effectiveness to offline scenarios when coupled with a clear limited-time justification. Notably, the framing of the message and the underlying motives for its uniqueness play a pivotal role in shaping the impact of scarcity messages(Jang et al., 2015; Roy & Sharma, 2015).

Interestingly, scarcity messages concerning limited quantity can impede purchase intention during the early stages of the purchasing journey, but exhibit a greater influence on purchase intentions during later stages (Cremer & Loebbecke, 2021). The timing of deploying scarcity messages holds significance, as they can potentially run counter to the overall sales strategy (Park et al., 2020; Zhao et al., 2021). The consumer shopping goal stage theory, as proposed by Lee and Ariely (2006)), posits that marketing activities aligned with consumers' shopping goals are more effective. In line with this notion, Sun et al. (2022) indicate that scarcity messages related to limited quantity prove more effective when consumers are at home, while scarcity messages tied to limited-time offers are more impactful when consumers are physically present at a store. Correspondingly, Ho and Lim (2018) present evidence that location-based and need-based matching personalization triggers unplanned purchases. Building upon this, we formulate the following hypothesis:

H2a: Scarcity messages are positively associated with need-based matching personalization of push notifications.
H2b: Scarcity messages are positively associated with location-based matching personalization of push notifications.

Price incentives are widely recognized and extensively employed marketing strategies for attracting consumers and stimulating purchase intentions (Chaudhuri et al., 2018). Specifically, price incentives enhance consumers' perception of available resources (Zhang et al., 2021). Price incentives encompass strategies such as reducing product prices or enhancing products with added features to make them more appealing, all without increasing the price (Zoellner & Schaefers, 2015). They play a pivotal role in bolstering consumers' preference for premium products (Miller et al., 2021) and enhancing brand and product recall (Crespo-Almendros & Del Barrio-García, 2014). However, it is imperative to exercise caution when employing price incentives, as, Haviv (2022) has demonstrated that their effectiveness in low-demand products is notably lower compared to high-demand products. This discrepancy can be attributed to the limited consumer searches for low-demand products in contrast to their high-demand counterparts.

To mitigate this issue, the implementation of need-based push notifications could prove beneficial by disseminating incentives for high-demand products in alignment with consumer needs. In light of the interconnected nature of consumers in today's digital age, it is important to recognize that consumers serve both as purchasers and information sharers (Sun et al., 2018). Consequently, targeting consumers based on their convenient location and timing can facilitate the sharing of price incentives within their online networks.
The consumer shopping goal stage theory posits that marketing activities that trigger consumers' shopping goals are more efficacious (Lee & Ariely, 2006). Therefore, price incentives that align with consumer goals or exhibit goal congruity hold greater influence (Roehm & Roehm, 2011). Additionally, Ho and Lim (2018) have demonstrated that location-based and need-based personalized approaches can evoke unplanned purchases.
In light of the aforementioned considerations, we hypothesize:

H3a: Price incentives are positively associated with need-based matching personalization of push notifications.
H3b: Price incentives are positively associated with location-based matching personalization of push notifications.

Theory of mood congruence

The principles of mood congruency theory suggest that individuals tend to recall events that align with their emotional state (Bower, 1981; Schwarz & Clore, 2003). Consequently, people in positive moods tend to remember pleasurable experiences, while those in negative moods are more likely to recall negative experiences. According to Ho and Lim (2018), one's emotional state can influence decision-making, guided by their current mood. While the terms mood, affect, and emotion are often used interchangeably, emotion and affect generally refer to specific temporary states, whereas mood encompasses a broader, more generalized emotional state. Therefore, in this study, we adopt the term "mood" to describe users' emotional states. Mood state-dependent memory posits that individuals access information in their memory primarily when they return to the same emotional state (Bower, 1981). For example, a person who was intoxicated will better recall experiences that occurred while they were drunk, and these memories become less accessible when they are sober.

Numerous studies have validated the influence of mood on users' decision-making across various domains, including cultural differences, tipping behavior, brand attitude, purchase intention, technological use or acceptance, and work environments. Willingness to pay and cultural norms impact mood transience; negative moods can increase willingness to pay more than positive moods(Maier et al., 2012). In contrast, collective cultures are more adaptable and prone to acclimatizing themselves to the prevailing environment compared to individualistic cultures. Building on the research by Ho and Lim (2018), emotional state or mood influences impulsive buying, and emotional extremity positively influences tipping behavior on live-streaming platforms(Chen et al., 2023). Positive moods alleviate privacy concerns related to disclosure behavior (Alashoor et al., 2022). In print advertising, positive moods positively influence brand attitudes (Batra & Stayman, 1990). Positive moods also enhance the technology acceptance (Djamasbi et al., 2010). Individuals in positive moods tend to utilize more informational cues, resulting in higher-quality decision-making (Djamasbi, 2007). Conversely, in organizational behavior, a positive mood may decrease the inclination for exploratory search, whereas a negative mood may reinforce persistence in continuing an exploratory search despite approaching deadlines (Knight, 2015). In a computerized work environment, employees are more likely to seek feedback from colleagues, especially when they perceive the feedback giver to be in a good mood (Ang et al., 1993). In this study, we propose that mood state could influence the relationship between push notifications and location and need-based personalization. Specifically, we hypothesize that individuals in positive moods would be more inclined to use decision aids and seek more information to support their online purchasing decisions. In contrast, individuals in negative moods may be less interested in receiving additional information to make online purchase decisions. In summary, promotional activities such as charity appeals, scarcity messages, and price incentives may be influenced by consumer mood, with the perception that mood changes based on location and personal needs. Therefore, we formulate the following hypotheses:

H4a: Consumer moods positively moderate the relationship between charity appeals and need-based matching of push notifications.
H4b: Consumer moods positively moderate the relationship between charity appeals and location-based matching of push notifications.
H4c: Consumer moods positively moderate the relationship between scarcity messages and need-based matching of push notifications.
H4d: Consumer moods positively moderate the relationship between scarcity messages and location-based matching of push notifications.
H4e: Consumer moods positively moderate the relationship between price incentives and need-based matching of push notifications.
H4c: Consumer moods positively moderate the relationship between price incentives and location-based matching of push notifications.

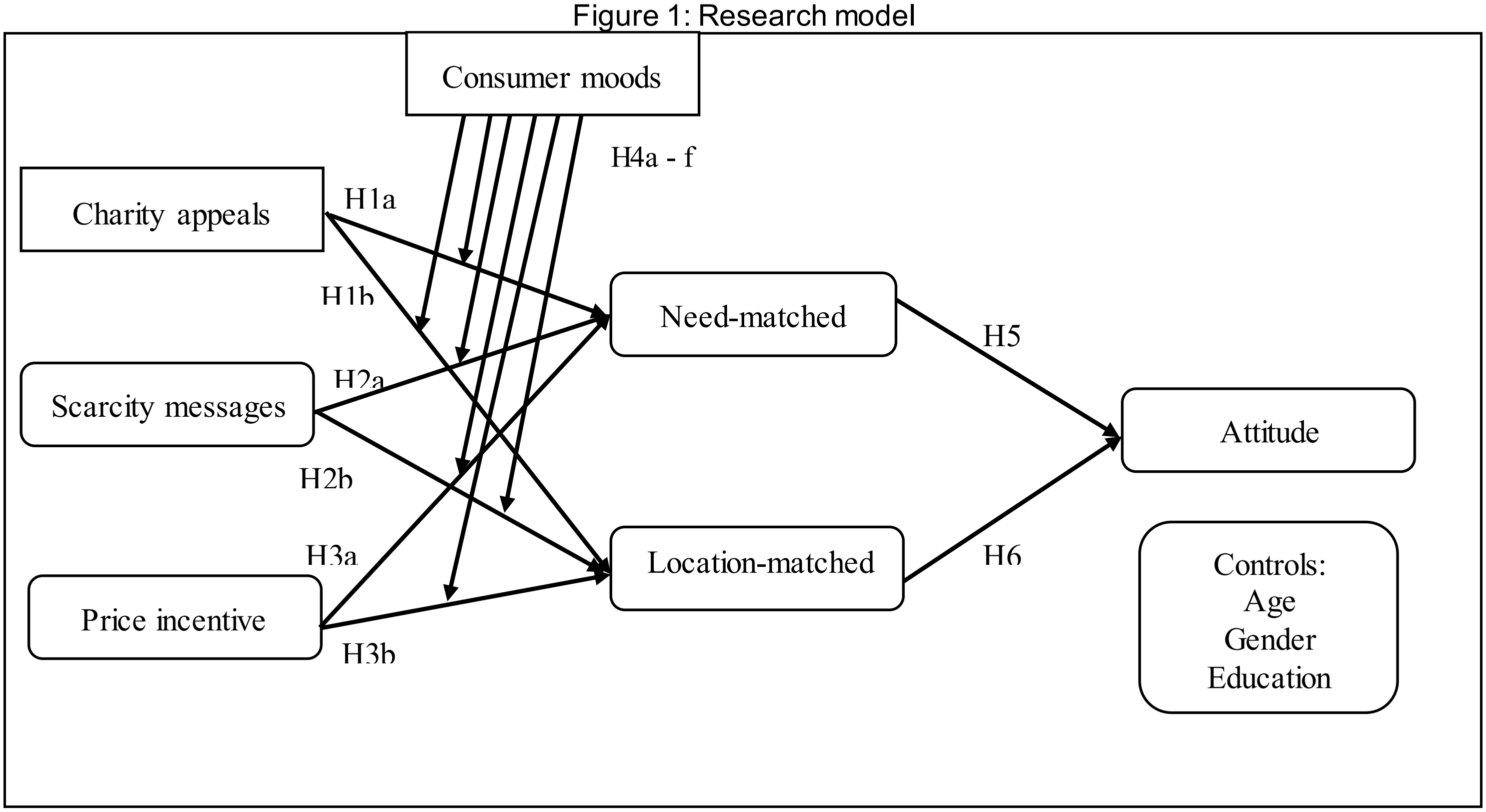
Needs are defined by the degree of essential well-being that something provides to a person, to the extent that they cannot do without it. Ho and Lim (2018) distinguish needs from personal wants or desires, describing them as distinctive requirements. Various types of needs exist, including emotional needs, e.g., Kansei (Xie et al., 2021), task-related needs, information needs, motivational needs, and physiological and psychological needs (Ho & Lim, 2018). Anunrojwong et al. (2022) conducted a study involving individuals with varying levels of needs and found that low-need individuals typically have alternative options, whereas high-need individuals do not. Consequently, targeting high-need individuals with push notifications addresses their requirements and enhances their attitude toward push notifications. Moreover, need-based systems are more effective for non-experts or individuals inexperienced in product assessment (Randall et al., 2007). Thus, need-based matching reduces push notification frustration and increases interest in receiving push notifications.

In contrast to need-based advertising, location-based advertising overcomes geographical barriers and connects with consumers at any time and place. Location-based advertising involves tailoring promotional activities to locations convenient for individuals to receive and engage with them (Bruner & Kumar, 2007). Consumers consider proximity an important and relevant aspect of their shopping goals (Molitor et al., 2020). Previous studies have shown that consumers prefer advertising that aligns with their locations (Ho & Lim, 2018). Today, the global positioning system makes implementing location-based advertising more accessible. Therefore, push notifications integrated with location-based matching are likely to be more appealing to consumers. The consumer shopping goal stage theory posits that marketing activities that trigger consumers' shopping goals are more effective (Lee & Ariely, 2006). In this context, proximity and the fulfillment of needs contribute to the successful achievement of consumers' shopping goals. Thus, we propose the following hypotheses:

H5: Need-based matching is positively associated with attitude toward push notifications acceptance and engagement.

H6: Location-based matching is positively associated with attitude toward push notifications acceptance and engagement.

Figure 1: Research model



METHODOLOGY

Research design

We adopted measures from prior studies to investigate the effect of personalized push notifications based on need-based and location-based matching. We slightly modified the question items to suit our context. The dependent variable is the attitude toward push notifications engagement and acceptance. We measured attitude using four question items. The question items include “Mobile notifications based on what I need and where I am are useful” and “Mobile notifications based on what I need and where I am are valuable”. A 7-point Likert scale to measure attitudes was used with 1=strongly disagree and 7 = strongly agree. Need-based matching was measured using four question items, adopted from (Ho & Lim, 2018). This construct consists of the following items “I prefer receiving mobile notifications when I need that particular product or service” and “I prefer receiving mobile notifications when I am thinking about that product or service”. A 7-point Likert scale was used to measure need-based matching, with 1= strongly disagree and 7= strongly agree. Location-based matching was measured using four question items adopted from (Xu et al., 2009). The question items include “With LBS, I am able to get up-to-date information whenever I need to” and “ With LBS, I am able to access relevant information or services at the right place”. Again a 7-point Likert scale was used to measure location-based matching with 1 = strongly disagree and 7 = strongly agree. Charity appeals were measured using four question items, adopted from (Barki & Hartwick, 1994; Erlandsson et al., 2018). The question items include “Charity appeals motivate me about a particular product” and “Charity appeals help me see the usefulness of a product”. A 7-point Likert scale was used to measure the items, with 1 = strongly disagree and 7 = strongly agree. Scarcity messages were measured using four question items adopted from (Aggarwal et al., 2011; Barki & Hartwick, 1994). The question items include “Scarcity messages motivate me about a particular product” and “Insufficient quantity notifications help me see the usefulness of product”. Price incentives were measured with four question items adopted from (Barki & Hartwick, 1994). The questions include “Price incentives motivate me about a particular product” and “Price discount promotion increase my intention to buy a particular product.” A 7-point Likert scale was used to measure price incentives and scarcity messages with 1 = strongly disagree and 7 = strongly agree.

Data collection and analysis procedure

For data collection, we developed an online survey using Google Forms, a widely employed platform for creating online surveys. Throughout the data collection process, we were meticulous in safeguarding the privacy of the respondents, and as such, we did not request any identifying information. The average duration for completing the survey was approximately 10 minutes. To ensure the attentiveness and engagement of participants, we strategically inserted attention trap questions at regular intervals within the survey. These attention trap questions guided respondents to select a specific choice or respond to a general question. The inclusion of three attention trap questions was aimed at enhancing the quality of the data collected, following the approach advocated by Oppenheimer et al. (2009). Participants who did not correctly answer the attention check questions were promptly excluded from the remainder of the survey. The final survey consisted of a total of 36 questionnaire items. Subsequently, we made the survey available on Amazon Mechanical Turk (mTurk). Amazon Mechanical Turk is a widely utilized platform for collecting data in social science studies, as attested by previous research (Buhrmester et al., 2016; Lowry et al., 2016).

FINDINGS

Demographic and sample size

We collected data from the Amazon Mechanical Turk (mTurk) crowdsourcing platform, a trusted and widely utilized resource in the realm of behavioral research, as attested by previous studies (Buhrmester et al., 2016; Lowry et al., 2016). The initial sample size comprised 293 responses. However, during the data-cleaning process, 13 responses were identified as outliers and subsequently removed. Consequently, the final dataset for analysis consisted of 280 responses. For comprehensive insights into our sample, please refer to Table 1 for demographic information and Table 2 for pertinent technical details about the study population

Table 1. Demographic information of sample size (n=280)

Age distribution	**No.**	**Gender**	**No.**	**Employment**	**No.**
Below 18	28	Male	160	Full-time	244
18 - 25	104	Female	120	Student	6
26 - 30	88			Unemployed	17
31 - 40	29			Other	13
41 - 50	31				
Total	**280**	**Total**	**280**	**Total**	**280**
Education	**No.**	**Nationality**	**No.**		
High school or below	35	American	197		
Vocational/technical school	22	Indian	57		
Undergraduate degree	185	Brazilian	13		
Postgraduate degree or higher	38	French	1		
		Other	12		
Total	**280**	**Total**	**280**		

Table 2 below displays the technical information about the data sample.

Table 2. Technical information of sample size (n=280)

Frequency of Mobile app use to shop	**No.**	**Monthly SMS usage**	**No.**	**Mobile app usage (Past six months)**	**No.**
Once a month	76	Less than 10 messages	22	Never	4
Twice a month	77	10 to 50 messages	90	Less than 10 times	40

Once a year	50	51 to 99 messages	88	10 – 29 times	98
2 - 4 times a year	28	100 to 300 messages	55	30 – 49 times	50
Once a week	49	More than 300 messages	25	50 times and above	88
Total	280	Total	280	Total	280

Pre-analysis and factorial validity

To assess and validate our model, we employed Smart PLS version 3.9.3 (Ringle et al., 2015). SmartPLS 3 is particularly well-suited for testing exploratory complex causal models.
Our data analysis proceeded through a two-step approach. In the initial step, we scrutinized the reliability and validity of the questionnaire items. This involved evaluating discriminant and convergent validity while addressing potential multicollinearity concerns, following the guidelines outlined, see (Gefen et al., 2000; Lowry & Gaskin, 2014). Subsequently, in the second step, we examined the relationships by employing bootstrapping. The measurement items adhered to established standards, with all factor loadings exceeding 0.7, and collinearity concerns were generally minimal, typically falling below 5 (Hair et al., 2011). Table 3 and Table 4 display the validity analysis and factor loading.

Table 3: Measurement Model Internal Consistency, Convergent, Discriminant Validity

	AT	CA	LBNM	LBPM	LM	NBNM	NBPM	NM	PI	SM	Alpha	CR	AVE
AT	**0.882**										0.904	0.904	0.777
CA	0.519	**0.935**									0.928	0.936	0.874
LBNM	0.361	0.682	**0.941**								0.936	0.938	0.886
LBPM	0.650	0.661	0.499	**0.911**							0.897	0.897	0.830
LM	0.750	0.527	0.374	0.572	**0.874**						0.897	0.899	0.764
NBNM	0.374	0.686	0.899	0.504	0.373	**0.944**					0.939	0.942	0.891
NBPM	0.694	0.589	0.415	0.861	0.616	0.405	**0.894**				0.875	0.875	0.800
NM	0.821	0.488	0.307	0.663	0.651	0.344	0.752	**0.887**			0.865	0.865	0.787
PI	0.535	0.506	0.343	0.568	0.577	0.318	0.584	0.541	**0.845**		0.798	0.806	0.713
SM	0.560	0.795	0.677	0.609	0.571	0.658	0.554	0.474	0.499	**0.895**	0.917	0.922	0.800

Note: AVE (Average variance), CR (Composite Reliability, rho a), Alpha (Cronbach's). Bolded are the square roots of AVEs. AT(Attitude), CA (Charity Appeals), LBNM (Location-based Negative Mood), LBPM (Location-based Positive Mood), LM (Location-based Matching), NBNM (Need-based Negative Mood), NBPM (Need-based Positive Mood), NM (Need-based Matching), PI (Price Incentive), SM (Scarcity Messages).

Table 4: Item Loadings and Cross Loadings from CFA

	AT	CA	LBNM	LBPM	LM	NBNM	NBPM	NM	PI	SM
AT1	**0.881**	0.396	0.249	0.562	0.668	0.256	0.636	0.714	0.480	0.442
AT2	**0.877**	0.470	0.312	0.590	0.660	0.329	0.610	0.742	0.475	0.489
AT3	**0.883**	0.439	0.317	0.540	0.666	0.308	0.598	0.719	0.448	0.496

AT4	**0.885**	0.526	0.397	0.600	0.649	0.426	0.606	0.720	0.485	0.548
CA2	0.520	**0.946**	0.663	0.639	0.522	0.669	0.568	0.474	0.467	0.772
CA3	0.438	**0.910**	0.627	0.572	0.438	0.628	0.515	0.408	0.435	0.735
CA4	0.493	**0.948**	0.624	0.639	0.513	0.627	0.565	0.481	0.513	0.725
LBNM1	0.323	0.621	**0.930**	0.463	0.338	0.815	0.373	0.276	0.317	0.618
LBNM2	0.360	0.664	**0.942**	0.477	0.370	0.857	0.419	0.298	0.329	0.651
LBNM3	0.336	0.640	**0.952**	0.467	0.347	0.865	0.379	0.293	0.323	0.642
LBPM1	0.603	0.594	0.424	**0.918**	0.520	0.446	0.782	0.603	0.505	0.532
LBPM2	0.589	0.613	0.475	**0.897**	0.527	0.465	0.786	0.584	0.536	0.569
LBPM3	0.585	0.599	0.464	**0.918**	0.515	0.467	0.786	0.625	0.512	0.564
LM1	0.659	0.544	0.387	0.545	**0.882**	0.396	0.566	0.574	0.525	0.534
LM2	0.662	0.457	0.323	0.505	**0.870**	0.317	0.569	0.619	0.465	0.485
LM3	0.669	0.455	0.317	0.536	**0.881**	0.331	0.576	0.568	0.551	0.516
LM4	0.629	0.381	0.276	0.404	**0.862**	0.253	0.434	0.512	0.474	0.457
NBNM1	0.336	0.667	0.842	0.491	0.348	**0.941**	0.388	0.306	0.275	0.609
NBNM2	0.340	0.662	0.841	0.492	0.344	**0.935**	0.389	0.323	0.316	0.635
NBNM3	0.381	0.617	0.862	0.448	0.364	**0.955**	0.371	0.344	0.306	0.620
NBPM1	0.600	0.488	0.360	0.756	0.533	0.362	**0.902**	0.669	0.537	0.483
NBPM2	0.621	0.550	0.392	0.779	0.562	0.383	**0.885**	0.690	0.505	0.516
NBPM3	0.642	0.541	0.362	0.775	0.557	0.340	**0.896**	0.657	0.525	0.488
NM1	0.720	0.391	0.236	0.580	0.618	0.266	0.688	**0.881**	0.472	0.381
NM2	0.730	0.505	0.334	0.625	0.520	0.379	0.648	**0.876**	0.456	0.459
NM4	0.736	0.403	0.248	0.561	0.593	0.273	0.665	**0.904**	0.512	0.421
PI2	0.495	0.355	0.171	0.453	0.514	0.165	0.500	0.510	**0.847**	0.359
PI3	0.391	0.466	0.396	0.503	0.450	0.343	0.481	0.397	**0.791**	0.451
PI4	0.462	0.473	0.325	0.491	0.495	0.313	0.498	0.455	**0.893**	0.465
SM1	0.521	0.698	0.574	0.555	0.541	0.538	0.518	0.430	0.458	**0.906**
SM2	0.534	0.746	0.591	0.558	0.550	0.548	0.530	0.467	0.469	**0.913**
SM3	0.475	0.733	0.650	0.591	0.479	0.661	0.495	0.403	0.431	**0.889**
SM4	0.469	0.667	0.616	0.473	0.466	0.624	0.434	0.389	0.427	**0.869**

Structural model specification

Our model prioritizes parsimony and comprises seven reflective first-order constructs. Among these constructs, "consumer moods" was meticulously formulated by considering both positive and negative emotions. To capture the multidimensional nature of consumer moods effectively, we employed a second-order construct approach.

Results

Table 3 below summarizes the results of the structural model. Importantly, the R^2 results for attitudes ($R^2 = 0.772$); location-based matching ($R^2 = 0.457$) and need-based matching ($R^2 = 0.431$). These are all substantial R square values, indicating higher predictive accuracy levels of our model (Hair et al., 2011; Henseler et al., 2009).

Moderation testing

Our central assumption regarding moderation posits that consumer moods will positively moderate the relationships between three distinct types of promotional activities: charity appeals, scarcity messages, and price incentives. To test this moderation effect, we employed a bootstrapping approach, consistent with methodologies used in prior studies (Hayes, 2009; Shrout & Bolger, 2002; Vance et al., 2015). However, the results yielded insignificance.
This lack of significance may be attributed to the intricacy of capturing genuine consumer mood through self-reported mechanisms, which necessitate participants to recall past experiences—potentially leading to inaccuracies in reporting. For a comprehensive presentation of the structural model results, please refer to Table 3 below.

Table 3. A Summary of Results from the structural model

Paths	Path Coefficient	t-statistics	p-value	Support?
Hypotheses				
Main effects				
H1a: Charity appeals → Need-based Matching	0.068	0.711	0.477	No
H1b: Charity appeals → location-based matching	0.052	0.354	0.723	No
H2a: Scarcity messages → Need-based Matching	0.111	1.102	0.905	No
H2b: Scarcity messages → location-based matching	0.283**	2.809	0.019	Yes
H3a: Price incentives → Need-based Matching	0.407***	4.570	0.000	Yes
H3c: Price incentives → location-based matching	0.413***	5.849	0.000	Yes
H5: Need-based Matching → Attitude	0.458***	4.649	0.000	Yes
H6: location-based matching → Attitude	0.436***	5.429	0.000	Yes
Moderations				

Table 3. A Summary of Results from the structural model				
Paths	**Path Coefficient**	**t-statistics**	**p-value**	**Support?**
Hypotheses				
H4a: Charity appeals → Consumer moods→ Need-based Matching	0.034	0.342	0.733	No
H4b: Charity appeals → Consumer moods→ location-based matching	0.057	0.057	0.515	No
H4c: Scarcity messages → Consumer moods→ Need-based Matching	0.036	0.419	0.675	No
H4d: Scarcity messages → Consumer moods→ location-based matching	-0.127	1.29	0.197	No
H4e: Price incentives → Consumer moods→ Need-based Matching	-0.044	0.548	0.346	No
H4f: Price incentives → Consumer moods→ location-based matching	-0.027	0.346	0.729	No
Control variables				
Age	-.059	-1.023	0.307	
Gender	-.044	-0.767	0.444	
Education	-.021	-0.36	0.719	

Notes: $^{***}p<.001$; $^{**}p<.01$; $^{*}p<.05$

DISCUSSION AND IMPLICATIONS

Discussions

The primary objective of this study was to investigate the impact of personalized push notifications on mobile shoppers engaged in online shopping. Our central hypothesis posited that personalized push notifications, based on both need-based and location-based matching, could mitigate the compulsive urge for constant checking, as identified by Iyer and Zhong (2021) and reduce passive acceptance of push notifications, as suggested by Chen et al. (2019). This growing sense of frustration arises from the significant amount of time users spend online, which can adversely affect personal productivity, as highlighted by Duke and Montag (2017). Our study sought answers to two pivotal questions: 1. To what extent does the personalization of push notification services from promotional activities (charity appeals, scarcity messages, and price incentives) influence mobile shoppers' attitudes towards engaging and accepting push notifications? 2. How does consumer mood moderate the relationship between promotional interventions (charity appeals, scarcity messages, and price incentives) and personalization (need-based and location-based matching) in online shopping?

Our findings have provided valuable insights into these questions. Specifically, Hypothesis 1a, which posited a positive association between charity appeals and need-based matching, was not supported. This suggests that charity appeals do not significantly influence the alignment of push notifications with users' specific needs. Hypothesis 1b, which suggested a link between charity appeals and location-based matching, was also not supported. This insignificance may be attributed to the nature of charity appeals, which often transcend geographical boundaries and may not depend on users' locations and their needs. Hypothesis 2a, proposing a positive association between scarcity messages and need-based matching, was not supported. Surprisingly, our results indicate no significant connection between scarcity messages and aligning notifications with users' specific needs. This discrepancy might be attributed to the context of online shopping, where time-scarcity promotions may require additional justifications to be as effective as their offline counterparts, as suggested by Hmurovic et al. (2023) argue that time-scarcity promotions which are effective offline might not be as effective online unless they are paired with some justifications.

Hypothesis 2b, which suggested a positive link between scarcity messages and location-based matching, received strong support. Our findings align with previous research Sun et al. (2022), emphasizing the role of both location and product type in the effectiveness of scarcity messages. Hypothesis 3a, proposing a positive connection between price incentives and need-based matching, was significantly supported. This suggests that price incentives effectively influence the alignment of push notifications with users' specific needs. Hypothesis 3b, suggesting a positive relationship between price incentives and location-based matching, was also supported. Our results indicate that price incentives effectively contribute to the alignment of notifications with users' locations. While our main effects were mostly significant and underscored the importance of our study, we acknowledge the limitations related to the complexity of capturing consumer mood using a self-reported mechanism. This may have contributed to the lack of statistical significance in the moderation effects. However, conceptually, we still argue that consumer mood may influence users' decisions based on findings from prior experimental studies. Thus, we cite this insignificance as a limitation of our study.

Theoretical and research implications

Our study makes significant contributions to existing theory in three key ways: first, push notification personalization**,** in the realm of personalization literature, our research stands out by focusing on push notifications and delving into the less explored negative aspects of this technology. While prior studies have investigated users' coping mechanisms in response to push notifications(Chen et al., 2019) and explored personalization in unplanned purchases (Ho et al., 2011; Ho & Lim, 2018). our study advances this field by proposing a novel approach. We suggest that tailoring push notifications based on need-based and location-based matching can effectively mitigate user frustration stemming from the inundation of generic push notifications. Secondly, addressing constant checking behavior**:** Our study responds to a dual narrative in the literature surrounding push notifications. While it is acknowledged that push notifications can reduce information anxiety, they have also been linked to an increase in constant checking behavior (Iyer & Zhong, 2022). In this context, our research pioneers efforts to understand and de-escalate this phenomenon of constant checking. By providing insights into how to manage push notification delivery timing effectively, we contribute to a better understanding of the factors influencing users' checking behavior. Thirdly, enhancing advertising strategies: In the broader domain of advertising literature, our study furnishes empirical evidence supporting the notion that the integration of personalized push notifications can serve as a potent tool for effective marketing interventions. Chen et al. (2019) highlighted that users do not dismiss push notification information as irrelevant; however, their passive acceptance often undermines its potential value. Our research underscores the significance of timing in launching these promotions. It underscores that the valuable information within push notifications can be harnessed effectively when delivered at the right moment.

Practical implications

Our study not only advances theory but also offers valuable insights that can bridge the gap between theory and practice, with implications in several key areas: Firstly, delicate push notification design: One of the significant practical contributions of our research lies in the proposition of a refined push notification design strategy. We urge mobile application designers to consider the potential benefits of personalized push notifications. In today's rapidly evolving technological landscape, where cutting-edge applications like global positioning systems (GPS) are commonplace, we anticipate the emergence of even more advanced technology capable of detecting personal needs and precise locations. Embracing such innovations will be pivotal for app developers and marketing practitioners seeking to create more effective online interventions. Secondly, enhancing productivity: mitigating the phenomenon of constant checking behavior, as explored in our study, has broader implications for individual and collective productivity. Reducing the compulsion for constant checking not only benefits individuals in terms of personal productivity but also holds the potential to enhance productivity across entire industries. By promoting healthier and more focused technology use, our findings align with the pursuit of increased efficiency in various professional contexts. Thirdly, targeted marketing strategies: Marketing professionals can harness the insights from our study to develop more strategic and impactful promotional activities. Our research highlights the effectiveness of promotional strategies grounded in both need-based and location-based matching. By tailoring marketing efforts to align with individual needs and geographic contexts, practitioners can optimize the success of their promotional campaigns.

Limitations

While our study contributes valuable insights, it is essential to acknowledge its limitations, which open doors for future research directions. Firstly, methodological limitation: Our study relied on self-reported mechanisms to gather responses, a common approach in survey research. However, the complexity of the consumer mood phenomenon suggests that alternative methods may be necessary to capture its nuances fully. Future studies could explore experimental designs or longitudinal approaches to provide a more comprehensive understanding of consumer mood. Secondly, sequential study design**:** To gain a deeper understanding of the Consumer Shopping Stage Goal Theory, we recognize the need for a more granular examination of the consumer shopping journey. Different promotional activities may be effective at distinct stages of the purchase process. Thus, future research could adopt a sequential study design to investigate the impact of personalized push notifications at various stages of the consumer journey, offering insights into the evolving dynamics of user behavior. Thirdly, measuring actual behavior**:** While our study focused on intentions, measuring actual consumer behavior is a critical avenue for future research. As noted by Erlandsson et al. (2018), attitudes do not always align with real-world actions. To bridge this gap, future studies could explore methods to observe and analyze actual user behavior in response to personalized push notifications, providing a more comprehensive picture of their impact. Lastly, exploring additional promotional activities: Our study concentrated on three specific promotional activities—charity appeals, scarcity messages, and price incentives. However, the ever-evolving online platform may offer numerous other promotional opportunities. Future research could expand its scope to investigate the effectiveness of alternative promotional strategies in the online environment. As highlighted by Hmurovic et al. (2023), strategies that succeed offline may not seamlessly translate to online platforms, making this an area ripe for exploration.

CONCLUSION

In conclusion, our study delved into the impact of personalized push notifications, based on need-based and location-based matching, on the attitudes of mobile shoppers toward push notification engagement and acceptance. This exploration was prompted by the growing frustration among users due to intrusive push notifications, leading to constant checking behaviors and diminished personal productivity. Our research offers compelling evidence on the potential of personalization to alleviate the intrusiveness of push notifications. Specifically, we uncovered that three distinct marketing strategies—charity appeals, scarcity messages, and price incentives—positively relate to personalization, encompassing both need-based and location-based matching. Furthermore, these marketing strategies contribute to enhancing consumers' attitudes, fostering greater willingness to engage with and accept push notifications. These findings not only advance the field of personalization but also offer actionable insights for practitioners seeking to craft more tailored and effective push notifications. In an era marked by information overload, our study provides guidance on leveraging personalization to enhance the user experience and engagement.

While our study did not yield evidence supporting the moderation hypothesis regarding the strengthening of the relationship between marketing strategies and personalization by consumer mood, it is essential to recognize the influence of emotional states on decision-making behavior, as demonstrated by prior research. We therefore advocate for further exploration of the role of consumer mood in the context of online shopping, urging researchers from diverse domains to contribute to this evolving prospect. In sum, our study contributes valuable insights to the intersection of personalization, marketing strategies, and consumer behavior in the realm of push notifications. It sets the stage for more nuanced and effective approaches to engage users while respecting their preferences and needs in the ever-evolving landscape of mobile commerce.

REFERENCES

Aggarwal, P., Jun, S. Y., & Huh, J. H. (2011). Scarcity messages. *Journal of Advertising*, *40*(3), 19-30.

Alashoor, T., Keil, M., Smith, H. J., & McConnell, A. R. (2022). Too Tired and in Too Good of a Mood to Worry About Privacy: Explaining the Privacy Paradox Through the Lens of Effort Level in Information Processing. *Information systems research*.

Ang, S., Cummings, L. L., Straub, D. W., & Earley, P. C. (1993). The effects of information technology and the perceived mood of the feedback giver on feedback seeking. *Information systems research*, *4*(3), 240-261.

Anunrojwong, J., Iyer, K., & Manshadi, V. (2022). Information Design for Congested Social Services: Optimal Need-Based Persuasion. *Management Science*. https://doi.org/10.1287/mnsc.2022.4548

Barki, H., & Hartwick, J. (1994). Measuring user participation, user involvement, and user attitude. *Mis Quarterly*, 59-82.

Batra, R., & Stayman, D. M. (1990). The role of mood in advertising effectiveness. *Journal of consumer research*, *17*(2), 203-214.
Bell, L., Garnett, C., Bao, Y., Cheng, Z., Qian, T., Perski, O., Potts, H. W., & Williamson, E. (2023). How Notifications Affect Engagement With a Behavior Change App: Results From a Micro-Randomized Trial. *JMIR mHealth and uHealth*, *11*(1), e38342.
Berger, V., & Burkhalter, L. (2022). Beyond one size fits all? An experimental study of the effects of stage-specific interventions to promote ecological online food shopping. *Journal of Consumer Behaviour*, *21*(5), 1040-1056.
Bies, S. M., Bronnenberg, B. J., & Gijsbrechts, E. (2021). How push messaging impacts consumer spending and reward redemption in store-loyalty programs. *International Journal of Research in Marketing*, *38*(4), 877-899.
Bower, G. H. (1981). Mood and memory. *American psychologist*, *36*(2), 129.
Bridges, E., & Florsheim, R. (2008). Hedonic and utilitarian shopping goals: The online experience. *Journal of Business Research*, *61*(4), 309-314.
Bruner, G. C., & Kumar, A. (2007). Attitude toward location-based advertising. *Journal of interactive advertising*, *7*(2), 3-15.
Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?
Chan, J. C., Jiang, Z., & Tan, B. C. (2009). Understanding online interruption-based advertising: Impacts of exposure timing, advertising intent, and brand image. *IEEE transactions on engineering management*, *57*(3), 365-379.
Chaudhuri, M., Calantone, R. J., Voorhees, C. M., & Cockrell, S. (2018). Disentangling the effects of promotion mix on new product sales: An examination of disaggregated drivers and the moderating effect of product class. *Journal of Business Research*, *90*, 286-294.
Chen, J. V., Tran, A., & Nguyen, T. (2019). Understanding the discontinuance behavior of mobile shoppers as a consequence of technostress: An application of the stress-coping theory. *Computers in Human Behavior*, *95*, 83-93.
Chen, X., Ji, L., Jiang, L., & Huang, J. T. (2023). The bright side of emotional extremity: Evidence from tipping in live streaming platform. *Information & Management*, *60*(1), 103726.
Cheng, Z., Li, K., & Teng, C.-I. (2022). Understanding the influence of privacy protection functions on continuance usage of push notification service. *Aslib Journal of Information Management*, *74*(2), 202-224. https://doi.org/10.1108/ajim-04-2021-0121
Choi, J., Rangan, P., & Singh, S. N. (2016). Do Cold Images Cause Cold-Heartedness? The Impact of Visual Stimuli on the Effectiveness of Negative Emotional Charity Appeals. *Journal of Advertising*, *45*(4), 417-426. https://doi.org/10.1080/00913367.2016.1185982
Cremer, S., & Loebbecke, C. (2021). Selling goods on e-commerce platforms: The impact of scarcity messages. *Electronic Commerce Research and Applications*, *47*, 101039.
Crespo-Almendros, E., & Del Barrio-García, S. (2014). The quality of internet-user recall: A comparative analysis by online sales-promotion types. *Journal of Advertising Research*, *54*(1), 56-70.
Djamasbi, S. (2007). Does positive affect influence the effective usage of a decision support system? *Decision Support Systems*, *43*(4), 1707-1717.
Djamasbi, S., Strong, D. M., & Dishaw, M. (2010). Affect and acceptance: Examining the effects of positive mood on the technology acceptance model. *Decision Support Systems*, *48*(2), 383-394.
Dogtiev, A. (2023). Push notifications statistics [Newsletter]. *Business of Apps*. https://www.businessofapps.com/marketplace/push-notifications/research/push-notifications-statistics/

Duke, É., & Montag, C. (2017). Smartphone addiction, daily interruptions and self-reported productivity. *Addictive behaviors reports*, *6*, 90-95.
Erlandsson, A., Nilsson, A., & Västfjäll, D. (2018). Attitudes and donation behavior when reading positive and negative charity appeals. *Journal of Nonprofit & Public Sector Marketing*, *30*(4), 444-474.
Fogel, J., & Setton, M. K. (2022). Advertising with scarcity messages and attitudes for luxury skin-care products. *Asia Pacific Journal of Marketing and Logistics*, *34*(10), 2285-2298.
Ford, J. B., & Merchant, A. (2010). Nostalgia Drives Donations. *The Power of Charitable Appeals Based on Emotions and Intentions*, *50*(4), 450-459. https://doi.org/10.2501/s0021849910091592
Gavilan, D., & Avello, M. (2023). Enabling smartphone push notifications: the effect of a framed opt-in request. *International Journal of Mobile Communications*, *21*(1), 1-18.
Gavilan, D., & Martinez-Navarro, G. (2022). Exploring user's experience of push notifications: a grounded theory approach. *Qualitative Market Research: An International Journal*, *25*(2), 233-255.
Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the association for information systems*, *4*(1), 7.
Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, *19*(2), 139-152.
Haviv, A. (2022). Consumer search, price promotions, and counter-cyclic pricing. *Marketing Science*, *41*(2), 294-314.
Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408-420.
HelpLama. (2023). What do consumers think about push notifications? [Newsletter]. *HelpLama*. https://helpdesk.helplama.com/what-do-consumers-think-about-push-notifications/
Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The use of partial least squares path modeling in international marketing. In *New challenges to international marketing*. Emerald Group Publishing Limited.
Hmurovic, J., Lamberton, C., & Goldsmith, K. (2023). Examining the Efficacy of Time Scarcity Marketing Promotions in Online Retail. *Journal of Marketing Research*, *60*(2), 299-328.
Ho, S. Y., Bodoff, D., & Tam, K. Y. (2011). Timing of adaptive web personalization and its effects on online consumer behavior. *Information systems research*, *22*(3), 660-679.
Ho, S. Y., & Lim, K. H. (2018). Nudging moods to induce unplanned purchases in imperfect mobile personalization contexts. *Mis Quarterly*, *42*(3), 757-778.
Hussain, A., Mkpojiogu, E. O., Almazini, H., & Almazini, H. (2017). Assessing the usability of Shazam mobile app. AIP Conference Proceedings,
Iyer, G., & Zhong, Z. (2022). Pushing notifications as dynamic information design. *Marketing Science*, *41*(1), 51-72.
Iyer, G., & Zhong, Z. Z. (2021). Pushing Notifications as Dynamic Information Design. *Marketing Science (in print)*.
Jang, W. E., Ko, Y. J., Morris, J. D., & Chang, Y. (2015). Scarcity message effects on consumption behavior: Limited edition product considerations. *Psychology & Marketing*, *32*(10), 989-1001.
Kamiya, A. S. M., & Branisso, D. S. P. (2021). In the right place at the right time: A review of mobile location-based marketing and a research agenda. *REMark*, *20*(2), 199.
Kim, S., & Childs, M. L. (2021). Passion for the past: Effect of charity appeals and nostalgia on clothing donation intentions. *Journal of Consumer Behaviour*, *20*(5), 1179-1190.
Knight, A. P. (2015). Mood at the midpoint: Affect and change in exploratory search over time in teams that face a deadline. *Organization science*, *26*(1), 99-118.

Kunkel, T., Hayduk, T., & Lock, D. (2023). Push it real good: the effects of push notifications promoting motivational affordances on consumer behavior in a gamified mobile app. *European Journal of Marketing*.
Ladeira, W. J., Lim, W. M., de Oliveira Santini, F., Rasul, T., Perin, M. G., & Altinay, L. (2023). A meta-analysis on the effects of product scarcity. *Psychology & Marketing*.
Lee, L., & Ariely, D. (2006). Shopping goals, goal concreteness, and conditional promotions. *Journal of consumer research*, *33*(1), 60-70.
Lee, Y., Seo, J. Y., & Yoon, S. (2020). Charity advertising: congruence between political orientation and cause of need. *International Journal of Advertising*, *39*(7), 943-962. https://doi.org/10.1080/02650487.2019.1710972
Liu, Y., Jiang, Z., & Choi, B. C. (2022). Pushing Yourself Harder: The Effects of Mobile Touch Modes on Users' Self-Regulation. *Information systems research*.
Lowry, P. B., D'Arcy, J., Hammer, B., & Moody, G. D. (2016). "Cargo Cult" science in traditional organization and information systems survey research: A case for using nontraditional methods of data collection, including Mechanical Turk and online panels. *The journal of strategic Information Systems*, *25*(3), 232-240.
Lowry, P. B., & Gaskin, J. (2014). Partial least squares (PLS) structural equation modeling (SEM) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE transactions on professional communication*, *57*(2), 123-146.
Maier, E., Wilken, R., Schneider, H., & Kelemci Schneider, G. (2012). In the mood to buy? Understanding the interplay of mood regulation and congruence in an international context. *Marketing Letters*, *23*, 1005-1018.
Majumdar, A., & Bose, I. (2018). My words for your pizza: An analysis of persuasive narratives in online crowdfunding. *Information & Management*, *55*(6), 781-794.
März, A., Lachner, M., Heumann, C. G., Schumann, J. H., & von Wangenheim, F. (2021). How you remind me! the influence of mobile push notifications on success rates in last-minute bidding. *Journal of Interactive Marketing*, *54*(1), 11-24.
Miller, C. J., Brannon, D. C., Salas, J., & Troncoza, M. (2021). Advertising, incentives, and the upsell: How advertising differentially moderates customer-vs. retailer-directed price incentives' impact on consumers' preferences for premium products. *Journal of the Academy of Marketing Science*, *49*(6), 1043-1064.
Molitor, D., Spann, M., Ghose, A., & Reichhart, P. (2020). Effectiveness of Location-Based Advertising and the Impact of Interface Design. *Journal of Management Information Systems*, *37*(2), 431-456. https://doi.org/10.1080/07421222.2020.1759922
Narang, U., & Shankar, V. (2019). Mobile app introduction and online and offline purchases and product returns. *Marketing Science*, *38*(5), 756-772.
Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*(4), 867-872.
Park, S., Rabinovich, E., Tang, C. S., & Yin, R. (2020). The impact of disclosing inventory-scarcity messages on sales in online retailing. *Journal of Operations Management*, *66*(5), 534-552.
Pielot, M., Church, K., & De Oliveira, R. (2014). An in-situ study of mobile phone notifications. Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services,
Randall, T., Terwiesch, C., & Ulrich, K. T. (2007). Research Note—User Design of Customized Products. *Marketing Science*, *26*(2), 268-280. https://doi.org/10.1287/mksc.1050.0116
Ringle, C. M., Wende, S., & Becker, J.-M. (2015). SmartPLS 3. SmartPLS GmbH. *Bönningstedt, Germany*.

Roehm, M. L., & Roehm, H. A. (2011). The influence of redemption time frame on responses to incentives. *Journal of the Academy of Marketing Science*, *39*, 363-375.

Roy, R., & Sharma, P. (2015). Scarcity appeal in advertising: Exploring the moderating roles of need for uniqueness and message framing. *Journal of Advertising*, *44*(4), 349-359.

Saputra, W., & Setiawan, E. B. Smart Application Development Recommendations for Mobile Food Traders Using Push Notification and Location Based Service (LBS) Based on Android.

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*(3-4), 296-303.

Septianto, F., & Tjiptono, F. (2019). The interactive effect of emotional appeals and past performance of a charity on the effectiveness of charitable advertising. *Journal of Retailing and Consumer Services*, *50*, 189-198. https://doi.org/https://doi.org/10.1016/j.jretconser.2019.05.013

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods*, *7*(4), 422.

Song, M., Noone, B. M., & Han, R. J. (2019). An examination of the role of booking lead time in consumers' reactions to online scarcity messages. *International Journal of Hospitality Management*, *77*, 483-491.

Song, T., Yi, C., & Huang, J. (2017). Whose recommendations do you follow? An investigation of tie strength, shopping stage, and deal scarcity. *Information & Management*, *54*(8), 1072-1083.

Sun, Q., Rajamma, R. K., Heisley, D. D., & Soliman, M. A. (2022). Examining scarcity in mobile promotion and purchase intention: the role of location. *Journal of Marketing Theory and Practice*, *30*(1), 1-19.

Sun, T., Viswanathan, S., Huang, N., & Zheleva, E. (2018). Designing promotional incentive to embrace social sharing: Evidence from field and online experiments. *Forthcoming at MIS Quarterly*.

Tam, K. Y., & Ho, S. Y. (2005). Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information systems research*, *16*(3), 271-291.

Tams, S., Legoux, R., & Léger, P.-M. (2018). Smartphone withdrawal creates stress: A moderated mediation model of nomophobia, social threat, and phone withdrawal context. *Computers in Human Behavior*, *81*, 1-9.

Tolin, D. F., Abramowitz, J. S., Brigidi, B. D., & Foa, E. B. (2003). Intolerance of uncertainty in obsessive-compulsive disorder. *Journal of anxiety disorders*, *17*(2), 233-242.

Trzebinski, W., Gaczek, P., & Marciniak, B. (2023). Is it better to communicate product information abstractly or concretely? The role of consumer product expertise and shopping-stage mindset. *Journal of Product & Brand Management*, *32*(2), 273-285.

Vance, A., Lowry, P. B., & Eggett, D. (2015). Increasing accountability through user-interface design artifacts. *Mis Quarterly*, *39*(2), 345-366.

Wohllebe, A., Hübner, D. S., Radtke, U., & Podruzsik, S. (2021). Mobile apps in retail: effect of push notification frequency on app user behavior. *Innovative Marketing*, *17*(2), 102-111.

Xie, N., Chen, D., Fan, Y., & Zhu, M. (2021). The acquisition method of the user's Kansei needs based on double matrix recommendation algorithm. *Journal of Intelligent & Fuzzy Systems*, *41*(2), 3809-3820.

Xu, H., Teo, H.-H., Tan, B. C., & Agarwal, R. (2009). The role of push-pull technology in privacy calculus: the case of location-based services. *Journal of Management Information Systems*, *26*(3), 135-174.

Zajonc, R. B., & Markus, H. (1982). Affective and cognitive factors in preferences. *Journal of consumer research*, *9*(2), 123-131.

Zhang, K., Cai, F., & Shi, Z. (2021). Do promotions make consumers more generous? The Impact of price promotions on consumers' donation behavior. *Journal of marketing*, *85*(3), 240-255.

Zhao, H., Wang, X., & Jiang, L. (2021). To purchase or to remove? Online shopping cart warning pop-up messages can polarize liking and purchase intention. *Journal of Business Research*, *132*, 813-836.

Zoellner, F., & Schaefers, T. (2015). Do price promotions help or hurt premium-product brands?: the impact of different price-promotion types on sales and brand perception. *Journal of Advertising Research*, *55*(3), 270-283.

DECISION SCIENCES INSTITUTE

The Impact of Process Degradation on Improvement Activity: An Exploratory Study

Brad C. Meyer
Drake University
Email: bradley.meyer@drake.edu

Dan Bumblauskas
Missouri Western State University
Email: dbumblauskas@missouriwestern.edu

Richard Keegan
Trinity College Dublin
Email: rkeegan@tcd.ie

Dali Zhang
Shanghai Jiao Tong University
Email: zhangdl@sjtu.edu.cn

ABSTRACT

This study explores the significance of process degradation to process improvement activity. Survey data was analysed from 351 improvement projects from 128 organizations in China, Ireland, and Iowa, across various industry sectors. Sixty-three percent of projects had processes that had degraded and 60% made at least one change to address degradation. The greatest driver of process degradation was demand growth. The percentage of changes that correct degradation was estimated at 28% by one measure and between 24% and 40% by another measure. Correction of degradation was an important driver of improvement in productivity and on-time delivery.

KEYWORDS: Process improvement, Process degradation, Improvement practice, Maintenance

INTRODUCTION

Years ago, one of the authors was introduced to short cycle kaizen while touring a manufacturing plant that had just completed its third kaizen event. The tour guide explained how, before the event, on one assembly line, the workload assigned to the various stations required one of the four workers to do less than 30% timewise of what was required at another station. The process improvement team balanced the line to reduce operator idleness. It produced a significant savings but begged the question: how did the line become this unbalanced? Surely it was not originally designed this way. The tour guide explained that several product design changes had occurred since the original formation of the line, but as the changes were implemented, no one readjusted the work assignments. In other words, through neglect, the process had become less efficient, and now, the correction finally made was being described as “continuous improvement.” But was it true improvement or simply maintenance?

Process improvement as a practice of an organization may take various forms. Industrial Engineering, TQM, Lean, and Kaizen are all examples. Within any of these forms, the specific changes made to processes or systems can be of two types: maintenance or enhancement.

Maintenance changes are corrective in nature, they restore something that has degraded. Maintenance on equipment usually corrects the wear and tear of equipment use. Belts become old and cracked, metal surfaces wear down, oil and lubricants are consumed or lost. A maintenance change on a process responds to degradation on multiple levels: degradation of the process elements, of the coordination between elements, of the synchronization between the process and the inputs and outputs, or of other alignments of the elements.

Enhancement changes improve the process state of the art. They usually add something new to the process, but sometimes simplify the process by a better design that requires fewer elements or relationships between elements. Enhancements are innovations, even if small, or are copies of an innovation developed elsewhere.

A company could show solid results from improvement activity, but still not grow in competitiveness if the improvements simply corrected process aspects that had deteriorated. Maintenance changes are critical and necessary, but enhancement changes push the process to new levels of performance. With this in view, an exploratory survey was conducted to make rough estimates of 1) the degree to which degradation is present in the processes addressed by improvement projects, 2) the proportion of changes made in such projects that serve to correct degradation, and 3) the degree to which maintenance changes are responsible for the gains reported from improvement projects. In other words, the research question is “how big of a role does process degradation play in the practice of process improvement?”

The first section below reviews the literature on process improvement to identify 16 types of changes made in a typical improvement project and nine focusing factors for improvement activity. The next section describes the survey instrument, followed by a section that details the survey process and the demographics of the respondents. A detailed presentation of the findings appears next followed by analysis and a discussion of the limitations of the study.

LITERATURE REVIEW

Process analysis and improvement have existed for as long as humans have performed tasks (Roser, 2016). Introspection into one’s activity is one of the hallmarks of humanness. The trajectory for modern forms of continuous improvement activity can be traced through such well-known figures as Taylor, the Gilbreths, Ford, Ohno, Shingo, Deming, Smith, Mikel, and Goldratt as well as through the work of lesser-known players such as Charles Babbage (Lewis, 2007), Florence Nightingale (Meyer and Bishop, 2007), Charles Allen (Huntzinger, 2002), Takuo Good (Robinson and Robinson, 1994) and Takishi Osada (Osada, 1991). Various authors have detailed the history of process improvement (Emerson and Naehring, 1988; Shingō and Robinson, 1990; Womack *et al.*, 1990; Schroeder and Robinson, 1991; Bhuiyan and Baghel, 2005; Schonberger, 2007; Roser, 2016). Continuous improvement is a non-negotiable in our dynamic, competitive global marketplace (Harrington, 1995; Kovach, Cudney and Elrod, 2011; Singh and Singh, 2012; Boer *et al.*, 2017).

Frederick Taylor spoke of “one best method” (Taylor, 2004) but actually Taylor and others after him considered the methods question as dynamic with a continual search for improvement. Taiichi Ohno explained in an interview, *“I told everyone that they weren’t earning their pay if they left the standardized work unchanged for a whole month.”* (Shimokawa & Fujimoto, 2009, pg 9).

Four broadly applied principles of process analysis and advancement are: 1: formalize process analysis and improvement (Taylor, 2004), 2: make use of process improvement experts (Emerson and Naehring, 1988), 3: leverage the knowledge of the worker (Gantt, 1901; Marcosson, 1945; Schroeder and Robinson, 1991), and 4: make use of the collective strength of improvement teams (Deming, 1986; Schroeder and Robinson, 1991; Liker, 1997; Fujimoto, 1999; Huntzinger, 2002; Goldratt and Cox, 2004).

The application of such principles leads to improvement activity in which processes are tangibly changed. Classic sources on process analysis and improvement yielded 16 improvement change types. These are listed below, grouped into 4 categories.

The first six are found in various industrial engineering and operations management writings (Niebel, 1967; Mogensen and Rausa, 1989; Huntzinger, 2002).

1. Add fixtures and tooling to improve quality, increase efficiency, and free up the operators’ hands.
2. Configure the workspace and process steps for efficient flow.
3. Locate tools and parts conveniently according to frequency of use.
4. Apportion tasks to operators in a way that balances workload.
5. Replace human effort with machine effort (including computers) to increase throughput and quality.
6. Change the interface between person and machine according to ergonomic considerations.

The next four improvement types apply to equipment degradation.

7. Repair equipment that is operating poorly.
8. Replace equipment that is operating poorly with identical equipment
9. Replace equipment that is operating poorly with newer equipment, same technology, but updated.
10. Replace equipment that is operating poorly with different, more advanced technology, hardware and/or software.

The next three improvement types were formalized and championed within the Toyota production system. (Womack et al., 1990; Osada, 1991).

11. Prevent careless mistakes that yield defects (poka yoke).
12. Remove items in the workspace that are not being used (“declutter”).
13. Add visual cues to the workspace to reduce thinking and search time.

The final three change types do not form a unified group but are advocated by various sources and/or surfaced as important in discussions with practitioners during the development of the data collection instrument. For example, one organization required each unit to make process improvements in four areas each quarter: productivity, quality, delivery, and safety. Safety related change (#16) was included because of this interaction.

14. Change the product design for more efficient processing.
15. Change capacity levels in accordance with demand needs and variation.
16. Add safety devices to protect the operator.

The formal deployment of improvement activity in an organization requires a structure. Such structures often have a focusing scheme, such as variance reduction or constraint management. Nine focusing perspectives were identified in common practice (Table 1). While popularized separately, these perspectives overlap in concepts and in practice. Many organizations blend various perspectives together, e.g., Lean Six Sigma.

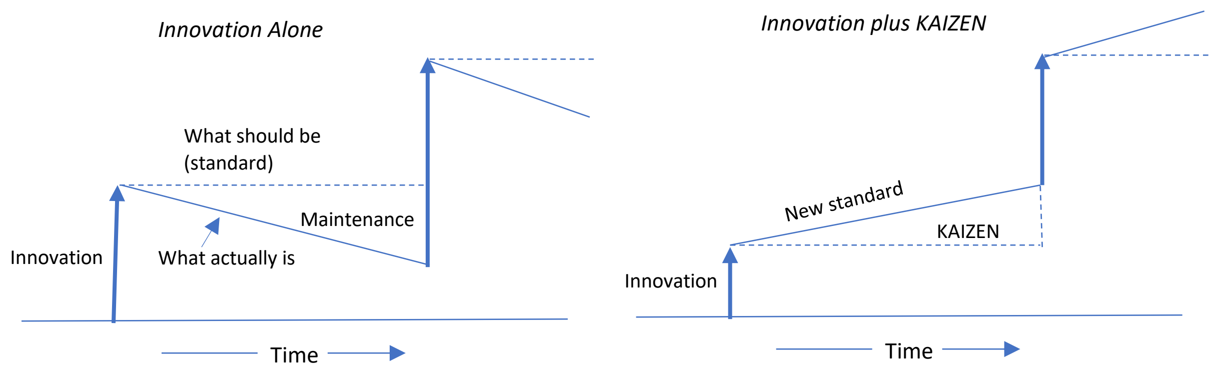
Table 1: Nine focusing perspectives for process improvement activity

Industrial Engineering	Focus on the efforts of a trained operational excellence professional. (Emerson and Naehring, 1988)
Total Quality Management	Focus on quality: reduce variation and rely on customer's definition of quality (Shewhart, 1926; Deming, 1986; Best and Neuhauser, 2006)
Six Sigma	TQM concepts with a more formalized managerial context including certified roles (black belts) and structured project selection practices. (Harry and Schroeder, 1999)
Lean	Focus on eliminating non-value-added effort (waste) and smooth even flow (Ohno, 1988; Kitaw and Belachew, 2007)
Kaizen (event)	Lean concepts implemented in short cycle projects (Laraia, Moody and Hall, 1999, Melnyk, et al 1998)
5S	Focus on workplace order and discipline (Osada, 1991; Hirano, 1995; Kobayashi, Fisher and Gapp, 2008)
Total Productive Maintenance	Focus on reliable and efficient operation of the equipment upon which production depends (Nakajima, 1988; Willmott and McCarthy, 2001; Ahuja and Khamba, 2008)
Benchmarking	Focus on discovering and imitating successful practices of other organizations (Camp, 1989; Voss, Åhlström and Blackmon, 1997)
Constraint Management	Focus on the bottlenecks that limit throughput (Goldratt and Cox, 2004)

The literature reports various surveys of process improvement practice. Surveys of lean practice (Yusuf and Adeleye, 2002; Doolen and Hacker, 2005; Nordin, Md Deros and Abd Wahab, 2010), TQM (Sila and Ebrahimpour, 2005; Khamalah and Lingaraj, 2007; Yusuf, Gunasekaran and Dan, 2007), and Six Sigma (Kumar, Antony and Douglas, 2009; Desai, Antony and Patel, 2012; Chakraborty and Chuan, 2013) examine the improvement program, rather than an individual improvement project. These surveys address the extent of application, the level of maturity, the effectiveness of various tools, the perceived success of CI methodologies, and other factors. What has not been undertaken is a survey of the types of changes made during an improvement event. There is also a lack of research identifying the state of project processes before the improvement activity commences.

Process improvement is ubiquitous in organizations, but so is process degradation. Meyer et. al. (2023) review the literature on process degradation to find several main threads of research in the literature. 1. Equipment degradation and maintenance, including TPM, 2. workspace disorganization and degradation as addressed by 5S, 3. process decay as a significant part of a process trajectory over time (see Figure 1), and 4. the challenge of maintaining the momentum of an improvement event or function over time and avoiding backsliding. They identify degradation as an increase in entropy over time and define workplace entropy as the degree of unbounded and equal access to non-productive (not adequately aligned) process states.

Figure 1: Degradation and Maintenance as Diagramed by Imai (1986), pg 26-27.



In the same paper, Meyer et. al. (2023) identify primary and secondary causes of entropy increase, i.e., process degradation. Primary causes include demand volume changes, incoming throughput changes, product design changes, wear and tear on processors, and behavioral inconsistency. Secondary causes are those that arise as the organization seeks to respond to primary causes. Scale expansion, scope expansion, equipment replacement and local reconfiguration are examples of secondary causes.

SURVEY INSTRUMENT DEVELOPMENT

The survey instrument was designed to gather detailed project information without being onerous to the respondent. Most questions could be answered from an A3 form or other project summary document. The core of this research is degradation and improvement activity. One question assessed degradation in the process before the improvement. That was question 2 in the final form. The specific changes that took place during or because of the project were the subject of question 4. Each of the 16 types of change identified in the literature was listed as an option in the question, check as many as apply.

Question 5 asked the respondent to estimate what percentage of the changes made were corrective in nature, as opposed to innovative. Shortcomings in the wording of this question are discussed later. Question 1 provided options for what prompted the project. Question 3 asked what types of deficiencies existed in the process before the improvement. The nine focusing perspectives were the subject of question 6 and the last question asked how performance changed in terms of productivity, quality, on-time delivery, safety, customization, and competitiveness.

The development of the survey proceeded through numerous steps. An initial survey draft was applied to project summary documents that were obtained from three different organizations.

After revisions, the survey form was sent to several practitioners. Their feedback led to further revision and a final feedback loop with other practitioners and academic colleagues. The final form can be found in the Appendix A.

The survey was distributed as described in the following section and each surveyed organization was asked to answer the questions for each of three recent improvement projects as well as to provide demographic data about their organization and their experience with process improvement.

RESPONDENTS AND DEMOGRAPHICS

To create a broad-based, exploratory dataset, a team of researchers from three regions was formed. The three regions were Europe (Ireland), North America (Iowa), and Asia (China). For each region, a team member obtained a list of organizations with continuous improvement activity. Surveys were emailed to Irish companies who had worked with Enterprise Ireland in its Lean Business Offer program (Keegan, 2011). Thirty-eight responses were received in response to a mailing of 650, for a response rate of 6%. In Iowa, 158 organizations that were members of the Iowa Lean Consortium (ILC, now part of the Iowa State University Center for Industrial Research and Service (CIRAS)) were contacted. Due to a small n from an initial mailing, a $30 Amazon gift card was offered as an incentive and ultimately 36 completed surveys were returned for a response rate of 23%. In China, interviews were conducted, instead of a mailed survey, with 54 of 91 organizations contacted, for a response rate of 59%. The interview process helped to ensure correct understanding of the survey questions which had been translated into Chinese. The Chinese organizations were located in Shanghai, Guangzhou, Shenzhen, Suzhou and other cities in the Yangtze River delta.

Though the response rates differed greatly, there was broad representation of industry sector, organization size, years of continuous improvement activity, and improvement program maturity in all three groups. While it cannot be proved that the relationships found in the data are broadly generalizable at the levels found, the data are useful for theory generation and as a proof of concept that such data can provide insight into process improvement practice. In total, the data describe 351 projects conducted by 123 organizations.

Manufacturing organizations constituted 44% of the organizations surveyed, and 56% were non-manufacturing. Thirty percent of the surveys were from Iowa, 28% from Ireland and 42% from China. Table 2 shows the breakdown for size of the organization and years of experience with continuous improvement. Table 3 shows the perspectives applied in the projects.

Table 2: Percentage of Surveys by Organization Size and Years of CI program

Employees	
under 20	10%
20 to 50	20%
50 to 100	16%
100 to 500	26%
500 to 2000	13%
over 2000	16%

Yrs doing CI	
1 to 2 years	29%
3 to 5 years	30%
6 to 10 years	22%
11 to 20 years	12%
over 20 years	7%

Table 3: Percentage Indicating a Given Perspective Was Applied

Perspective	% circled
Lean	61%
A kaizen event	43%
5S	35%
TPM	9%
Six Sigma	17%
TQM	10%
Bottleneck or constraint management (TOC)	22%
Benchmarking	16%
Industrial engineering (IE)	20%
Other	6%

FINDINGS

Process Degradation and Deficiency

Degradation was a sizable factor in project processes (Figure 2). Almost two thirds of the processes were reported as having degraded in performance, some from multiple contributing factors. The largest individual cause was demand growth, reducing performance in 39% of project processes, followed by product changes at 20%, changes that affected integration of the process at 19%, and wear and tear at 7%. Only in 21% of the projects was the process "working as well as ever". Some processes had capacity issues without degradation, making growth a concern in 61% of processes. Figure 3 breaks this out by demographic category.

Figure 2: 95% Confidence Interval for Percent Reporting Degradation Cause as Project Began

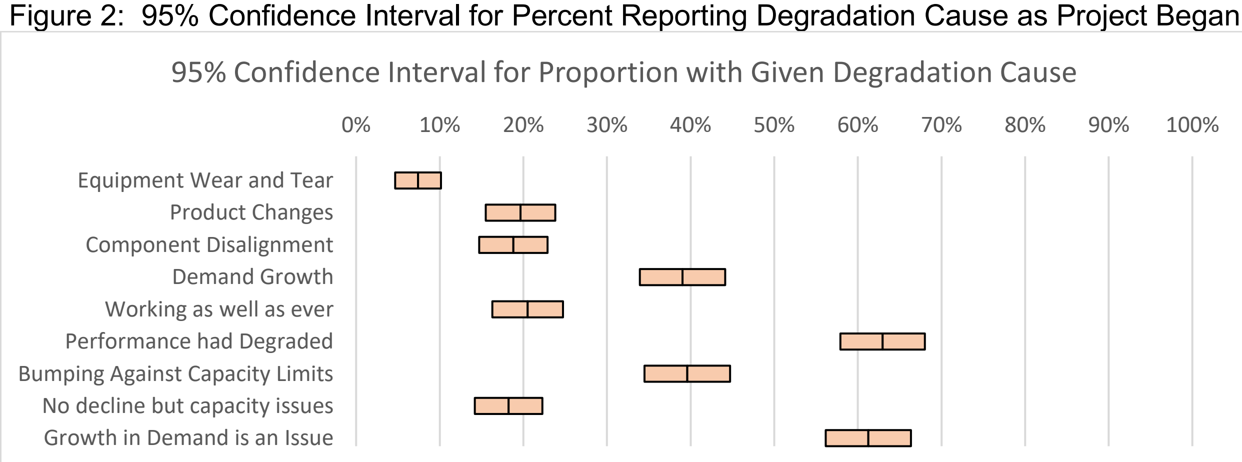


Figure 3: Average Percent Reporting Degradation Cause By Demographic Trait

	Overall	Ireland	Iowa	China	small	medium	large
Equipment Wear and Tear	7%	11%	7%	5%	5%	10%	7%
Product Changes	20%	24%	14%	21%	15%	24%	19%
Component Disalignment	19%	28%	18%	14%	16%	22%	18%
Demand Growth	39%	46%	48%	28%	42%	41%	34%
Working as Well as Ever	21%	8%	17%	31%	23%	19%	20%
Performance Had Degraded	63%	70%	66%	56%	59%	67%	60%
Bumping Against Capacity Limits	40%	52%	53%	22%	28%	44%	45%
No Decline But Capacity Issues	18%	20%	21%	16%	17%	19%	19%
Growth in Demand Is an Issue	61%	72%	74%	45%	63%	63%	56%

	Mftg	Non-mftg	1 to 2 yrs	3 to 5	6 to 10	11 to 20	over 20
Equipment Wear and Tear	12%	4%	14%	4%	5%	7%	4%
Product Changes	26%	15%	20%	26%	15%	19%	4%
Component Disalignment	22%	16%	25%	10%	22%	26%	9%
Demand Growth	40%	38%	35%	48%	35%	36%	35%
Working as Well as Ever	12%	27%	15%	22%	23%	24%	26%
Performance Had Declined	68%	59%	64%	70%	55%	67%	48%
Bumping Against Capacity Limits	46%	35%	37%	41%	36%	48%	43%
No Decline But Capacity Issues	21%	16%	22%	13%	23%	12%	22%
Growth in Demand Is an Issue	64%	59%	61%	65%	60%	57%	57%

Figure 4 provides perspective on degradation. While only 63% of processes had degraded over time, 94% of processes had some form of deficiency. The listed deficiencies from question 3 of the survey could have resulted from degradation or they could have existed in the process since the original implementation. Figure 5 breaks this down by demographic category.

Figure 4: 95% Confidence Interval for Percent Reporting Deficiency Type as Project Began

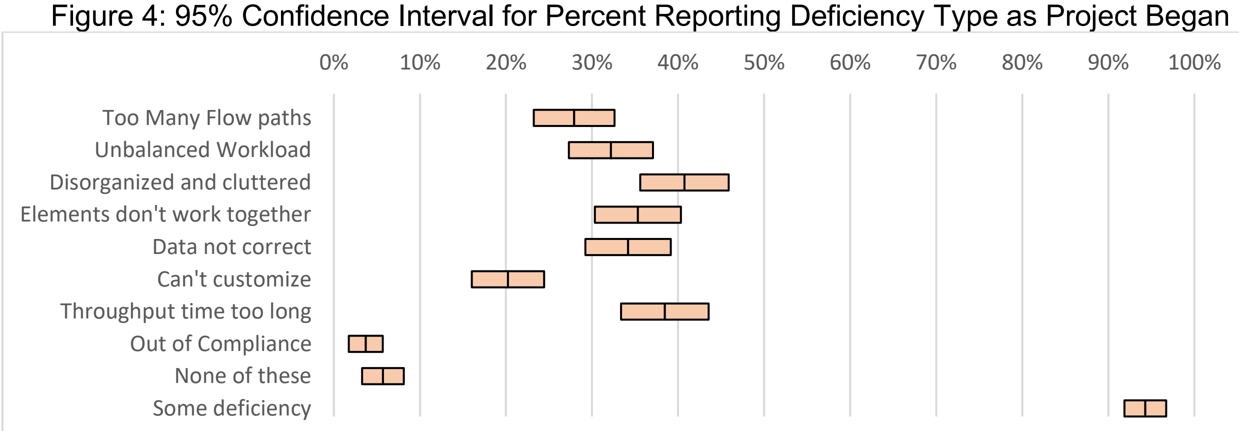


Figure 5: Average Percent Reporting Deficiency Type by Demographic Trait

	Overall	Ireland	Iowa	China	small	medium	large
Too Many Flow Paths	28%	42%	43%	7%	17%	34%	31%
Unbalanced Workload	32%	33%	46%	22%	18%	31%	48%
Disorganized and Cluttered	41%	43%	44%	36%	46%	42%	34%
Elements Don't Work Together	35%	34%	42%	32%	25%	35%	47%
Data Not Correct	34%	35%	29%	37%	40%	36%	26%
Can't Customize	20%	4%	11%	37%	31%	14%	18%
Throughput Time Too Long	38%	49%	62%	14%	25%	42%	47%
Out of Compliance	4%	7%	2%	3%	7%	2%	3%
None of These	6%	4%	5%	7%	7%	5%	6%
Some Deficiency	92%	94%	94%	90%	89%	95%	92%

	Mftg	Non-mftg	1 to 2 yrs	3 to 5	6 to 10	11 to 20	over 20
Equipment Wear and Tear	35%	23%	30%	35%	18%	29%	17%
Product Changes	44%	23%	34%	36%	23%	40%	22%
Component Disalignment	44%	38%	38%	46%	44%	31%	35%
Capacity Limits	39%	32%	30%	36%	32%	43%	52%
Demand Growth	22%	44%	31%	41%	33%	24%	39%
Working as Well as Ever	13%	26%	19%	18%	27%	24%	9%
Performance Had Declined	44%	34%	37%	41%	37%	40%	35%
Growth in Demand Is an Issue	3%	5%	1%	6%	4%	2%	9%
No Decline But Capacity Issues	4%	7%	5%	4%	8%	12%	0%
Some Deficiency	94%	91%	95%	92%	88%	88%	100%

Changes Made

Figure 6 shows changes made by the project. Changed work assignments, reorganizing the workplace, and reconfiguring flow were the most common changes made overall. Replacing a piece of equipment with an identical item was the least common while almost a third of the projects replaced equipment with more advanced technology. Note that changed work assignments would be a common side effect of other changes and not necessarily a rebalancing of workload. Figure 7 breaks this out by demographic category.

Figure 6: 95% Confidence Interval for Percent Reporting This Change

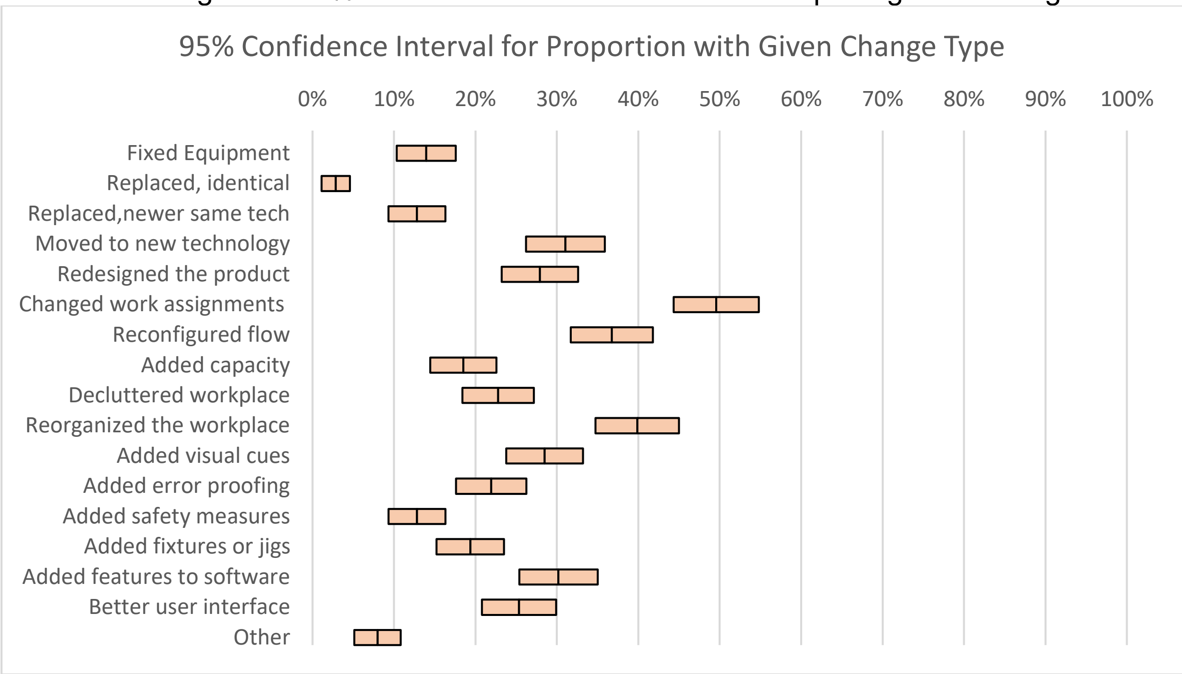


Figure 7: Average Percent Reporting Given Change Type by Demographic Category

	Overall	Ireland	Iowa	China	small	medium	large
Fixed Equipment	14%	27%	15%	5%	7%	18%	16%
Replaced, Identical	3%	2%	2%	4%	1%	4%	3%
Replaced, Newer Same Tech	13%	20%	8%	12%	11%	19%	6%
Moved to New Technology	31%	31%	37%	27%	27%	31%	35%
Redesigned the Product	28%	22%	14%	42%	36%	25%	24%
Changed Work Assignments	50%	49%	58%	44%	50%	41%	60%
Reconfigured Flow	37%	45%	50%	22%	26%	41%	42%
Added Capacity	19%	14%	18%	22%	18%	20%	16%
Decluttered Workplace	23%	31%	26%	15%	22%	24%	21%
Reorganized Workplace	40%	59%	45%	24%	37%	43%	39%
Added Visual Clues	28%	37%	25%	25%	27%	29%	29%
Added Error Proofing	22%	24%	34%	12%	13%	22%	32%
Added Safety Measures	13%	10%	12%	15%	11%	14%	13%
Added Fixtures or Jigs	19%	31%	14%	16%	14%	24%	18%
Add Features to Software	30%	26%	21%	40%	41%	25%	27%
Better User Interface	25%	16%	10%	42%	42%	20%	16%
Other	8%	9%	16%	1%	4%	10%	10%

	Mftg	Non-mftg	1 to 2 yrs	3 to 5	6 to 10	11 to 20	over 20
Fixed Equipment	25%	5%	16%	12%	13%	19%	9%
Replaced, Identical	4%	2%	3%	0%	4%	7%	4%
Replaced, Newer Same Tech	17%	9%	15%	11%	12%	14%	13%
Moved to New Technology	33%	29%	24%	36%	31%	40%	26%
Redesigned the Product	22%	32%	29%	27%	29%	29%	17%
Changed Work Assignments	45%	53%	41%	57%	47%	48%	65%
Reconfigured Flow	45%	30%	38%	40%	32%	38%	30%
Added Capacity	16%	21%	11%	23%	23%	14%	26%
Decluttered Workplace	31%	16%	28%	24%	19%	19%	13%
Reorganized Workplace	53%	30%	43%	37%	42%	33%	43%
Added Visual Clues	28%	29%	33%	29%	27%	21%	22%
Added Error Proofing	26%	19%	19%	25%	21%	36%	0%
Added Safety Measures	13%	13%	11%	10%	18%	12%	17%
Added Fixtures or Jigs	24%	15%	14%	21%	24%	17%	26%
Add Features to Software	21%	38%	29%	36%	29%	21%	26%
Better User Interface	9%	38%	22%	33%	26%	12%	30%
Other	11%	6%	9%	5%	9%	14%	4%

Changes were classified into two types: a maintenance change that restored a previous level of performance or an enhancement change that improved the state of the art of the process. The classification required judgement due to lack of detail about the specific changes made and the operating policies and technology before the project occurred. Maintenance changes included fixing equipment, replacing equipment with like equipment, adding capacity, and decluttering the workplace. Enhancement changes include adding fixtures or jigs, adding new software features, replacing with better technology, changing to a new user interface, adding error proofing devices, adding visual cues, redesigning the product, and adding safety devices. There were five change types of such a nature that could have been a correction or could have been a movement to something new: reconfiguring flow, reorganizing the workplace, making assignment changes, removing unnecessary items from the workplace, and other. For those cases, the results were separated based on whether the process was reported as having degraded in a relevant way. If so, a change was categorized as maintenance, if not, it was categorized as enhancement. The classifications are shown in Table 4.

Table 4: Categorization of Changes as Maintenance or Enhancement

Change	Categorization	Rationale
added fixtures or jigs to make the task more efficient	enhancement	This might address a deficiency, but is not a return to a previous condition
moved things around physically or in sequence of steps to improve the flow	maintenance enhancement	If the process was described in an earlier question as degraded and having flow problems If the process did not originally have flow problems
reorganized the workplace	maintenance enhancement	If the process was described in an earlier question as degraded and disorganized If the process had not degraded or was not disorganized
made work assignment changes	maintenance enhancement	If the process was described in an earlier question as having degraded and with unbalanced loads If the process had not degraded
added some new features to the software to make the task more efficient	enhancement	New features bring about new options, so not restorative.
fixed various problems with the equipment	maintenance	Corrects problems and restores to working order
replaced equipment with basically identical equipment	maintenance	This is a routine action when a machine can no longer be economically repaired
replaced equipment with newer equipment, same technology, but better	enhancement	This has some elements of maintenance, but because it is an upgraded version, it is classified as advancement
changed to a new technology, a different way of processing	enhancement	This moves the process to new levels of performance.
changed to software with a better (more efficient) user interface	enhancement	This is new technology in the software space.
put error proofing devices in place (to protect the product)	enhancement	This does not restore a process to a former condition but adds something new.
removed unnecessary equipment, material, and tools from the workplace/process	maintenance enhancement	If the process was described as having degraded and workplace was cluttered If the process had not degraded or was not cluttered
added signage or visual control elements	enhancement	This does not restore to a previous condition but adds something new.
redesigned the product itself, to better the process	enhancement	This is not a restorative change; it is something new.
added capacity	maintenance	Does not add anything new, just more of the same, restores the degradation that occurred from lack of capacity.
put safety devices in place (to protect the employee)	enhancement	Does not restore a previous state, but adds something new to the process
other	maintenance enhancement	If performance had declined If performance had not declined

Improvement Effort Devoted to Maintenance Changes

The degree to which improvement effort was devoted to correcting degradation was estimated in two ways: from the count of the change types made and by approximation from the respondents' estimate of percent corrective changes.

From count of the change types made

Table 5 reveals that maintenance was an important aspect of improvement projects, while not the dominant aspect. Sixty percent of projects included at least one maintenance change and 96% of projects made at least one enhancement change. Note, this is a rough figure since what is counted is the number of boxes checked for the two change categories. A project that repaired equipment might have repaired four different aspects of the machinery or might have repaired two different machines, but that would count as one change type in these numbers. The average number of different maintenance change types per project was 1.1. The average number of enhancement change types was 2.8. Thus 28% of counted change types were maintenance changes. Figure 8 breaks this out by demographic category. Figure 9 was added to show an important result. Each respondent was asked to characterize the maturity of their CI program. One of the options was "in danger of losing management support." It is noteworthy to find that for those who gave that response, the ratio of enhancement changes to maintenance changes was much lower than for the other options. CI programs fall out of favor when they are focused on fixing what is broken instead of advancing the state of the art.

Table 5: Percent Reporting and Average Number of Maintenance and Enhancement Changes

Any maintenance changes?	60%
Any enhancement changes?	96%
Ave. number of maintenance change types	1.1
Ave. number of enhancement change type	2.8

Figure 8: Number of Change Types by Demographic Category

	Overall	Ireland	Iowa	China	small	medium	large
maintenance changes	1.1	1.5	1.4	0.6	0.9	0.8	1.8
enhancement changes	2.8	2.9	2.5	3.0	2.9	2.6	4.0

	Mftg	Non-mftg	1 to 2 yrs	3 to 5	6 to 10	11 to 20	over 20
maintenance changes	1.4	0.8	1.1	1.2	1.1	1.1	0.7
enhancement changes	2.6	2.9	2.7	2.9	2.9	2.7	2.8

Figure 9: Number of Change Types by Characterization of CI Program

	young, growing	ingrained, healthy	ingrained, boring	in danger with mgmt	other
maintenance changes	1.3	1.0	0.9	1.3	0.4
enhancement changes	2.7	3.0	2.9	1.9	1.9
ratio: ehance/main	2.0	2.9	3.3	1.4	4.4

From estimated percentage of corrective changes

Question 5 asked the respondents to estimate the percentage of changes in the project that were *corrective or standard* and percentage considered *creative or novel*. The average value for "corrective or standard" was 61% and the average value for creative or novel was 39%. It could be argued that there are ways to address degradation that are creative and novel, and that the word "standard" could describe changes that are not corrective. To estimate the percentage of corrective changes, it was assumed that these two roughly cancel each other out and thus approximately 60% of changes are corrective. This value was corroborated by a question in the demographics section that asked respondents to rate the focus of their improvement program on a scale from problem solving (0) to innovation (7). The average score was 2.54, corresponding to a 64% focus on problem solving and 36% on innovation.

Corrective changes include both those that address degradation and those that address other deficiencies in the process that did not necessarily arise from deterioration. Figures 2 and 3 indicated that 63% of projects had degradation issues whereas 94% of projects had deficiencies calling for correction. These values can be used to set an upper and lower bound on the percentage of maintenance changes. If the 63% included no non-degradation deficiencies, then 63/94 of the 60% of correction changes would be maintenance changes, or 40%. Conversely, if all 63% of the processes with degradation issues also had other non-degradation deficiencies that were addressed by corrective changes, then 63/(63+94) of the 60% would be maintenance changes, or 24%. Thus, it is estimated that between 24% and 40% of changes made in the projects were changes to correct degradation. This is consistent with the earlier estimate of 28%.

Level of Improvement Seen from Maintenance Changes

The impact of degradation can be measured by the degree of improvement accomplished by changes that correct degradation. When a degraded process is restored and performance jumps up dramatically, it is clear that the organization had been missing out on productivity, on-time delivery, and/or some other performance measures before the maintenance change occurred. The average number of maintenance and enhancement changes for the four levels of improvement in each of the six performance areas, from question 6, is shown in Figure 10. Note, for example, more significant improvement in productivity corresponds to a higher number of maintenance change types, while the number of enhancement change types does not correspond to more significant gains in productivity. This indicates that enhancement changes were not driving productivity improvement. Competitiveness, however, sees more improvement with increased amounts of both kinds of change types.

Figure 10: Average Maintenance and Advancement Changes by Level of Improvement and Performance Area

	Productivity		Quality		On-time Delivery		Safety		Customization		Competitiveness	
	Main.	Adv.	Main.	Adv.	Main.	Adv.	Main.	Adv.	Main.	Adv.	Main.	Adv.
change for the worse	0.00	2.00	NA	NA	NA	NA	0.76	2.56	1.00	2.00	NA	NA
no change	0.54	2.92	0.54	2.42	0.67	2.81	1.44	2.93	1.10	2.47	0.65	2.23
a little improvement	0.89	2.74	1.19	2.72	1.05	2.64	1.70	3.60	1.12	2.98	1.02	2.80
significant improvement	1.61	2.78	1.48	3.32	1.66	2.97	1.00	1.00	0.96	3.30	1.49	3.15

Because the degree of improvement was surveyed as an ordinal variable, ordered multinomial regression was used to determine the impact of each change on the six types of improvement. For each of the improvement areas, a best fit model was determined by minimizing AIC, the Akaike information criterion, while verifying acceptable goodness of fit with the Hosmer and Lemeshow test (Hosmer and Lemesbow, 1980) and the Lipsitz test (Lipsitz, Fitzmaurice and Molenberghs, 1996). In the tables that follow, the change types that were significant in the best fit model are grouped into maintenance changes and enhancement changes. For each change, the percentage of projects that contained that change, the odds ratio, including a 95% confidence interval, and the p value of the variable in the regression model are shown in the table. The odds ratio can be interpreted as follows: the value shown describes how much more likely it is that the project will be classified as having significant change than it will be described as having no change or a little change. For example, in Table 6, projects that fixed various problems with the equipment will be 2.8 times more likely to see significant change in productivity than those projects that did not fix various problems with the equipment. Only change types whose coefficients were significant at the 0.10 level are shown.

Table 6: Odds Ratios for Significant Variables Predicting Improvement in Productivity

Type	Change	% of projects	**Odds Ratio**	OR 2.5%	OR 97.5%	p value
main	fixed equipment	14%	**2.8**	1.5	5.4	0.002
main	reconfigured flow - maintenance	14%	**2.6**	1.3	5.5	0.010
main	reorganized the workplace - maintenance	19%	**2.2**	1.0	4.8	0.040
main	other - maintenance	5%	**2.4**	0.9	6.5	0.075
enh	moved to new technology	31%	**2.2**	1.4	3.5	0.001
enh	reorganized the workplace - enhancement	21%	**2.2**	1.3	3.9	0.004
enh	added safety measures	13%	**0.4**	0.2	0.7	0.003
enh	better user interface	25%	**0.5**	0.3	0.8	0.004

Eight change types influenced productivity improvement in the best fit model, four were maintenance changes and four were enhancement changes. Two of the enhancement changes had odds ratios less than 1, meaning when those changes occurred, it was more likely that there was little or no improvement in productivity. All the significant maintenance changes increased the odds that the project had significant improvement in productivity.

Odds ratio tables for the other surveyed performance areas are found in Appendix B. Maintenance improvements seemed to play a larger role in on time delivery, but a more balanced role for quality, safety, and competitiveness. No maintenance change variables were significant to the ability to provide a customized product.

ANALYSIS AND IMPLICATIONS

What part does degradation play in improvement activity? Sixty-three percent of projects in the data set experienced decline in performance and 60% reported making at least one change that could be characterized as process maintenance. Twenty-eight percent of improvement change types were process maintenance and it is estimated that between 24% and 40% of process changes address degradation. In addition, improved productivity and on-time delivery are more likely to be associated with maintenance changes than with enhancement changes. Maintenance changes are also strong contributors, balanced with enhancement changes, to improvement in quality and competitiveness.

The loss in performance due to degradation is significant enough to warrant attention. Management should be alert to ways to prevent and reduce degradation at the process level. A process where the workloads have become unbalanced over time could be corrected by a reallocation of tasks to workers. But the imbalance could be prevented by employing a U-shaped line and flexible work responsibilities so that it becomes impossible for the work to become unbalanced. Concepts borrowed from equipment maintenance, such as predictive maintenance, can be applied at the process level which may require innovative process sensor mechanisms. For example, an early signal that degradation has occurred is when a person performing a process creates a work-around (Holweg, Staats and Upton, 2018). The new work-around behavior is a signal that there has been a change in the process that has disrupted the alignment. This signal could be captured by a 5S-like audit of the method, a daily standup scrum discussion, or even by an computer implemented algorithm that mines user activity. (*Process Mining and Execution Management Software | Celonis*, no date).

The survey results confirm the observation made by (Holweg *et al.*, 2018) who explain that one of the main drivers of operations entropy is growth.

"All the good things that you'd like to see happen to a company—more customers, growth in sales, new products, new locations—contribute to increasing the variation and complexity with which the operation has to cope… the irony of this operational entropy, this increasing disorder is inescapable. In the growth that a company longs for lie the seeds for the very operations headaches that can stall that growth." pgs. 167-8.

This provides direction for future research in restraining degradation: learn to manage growth in a way that minimizes the deterioration of the process. A starting point for such a study is in the ideas of Taiichi Ohno and the Toyota production system, which was designed in such a way as to work well under low volume and changes in demand (Shimokawa and Fujimoto, 2009). Quick set ups, flexible production lines, defect prevention, and Kanban inventory control, for example, are techniques that support volume scalability.

Another driver of process degradation is change to the process. But more specifically, it is the randomness of changes that occur, i.e., each change may align well with some elements of the process, but not well with other elements of the process. When product design is altered, suppliers are changed, products are added or dropped, equipment is replaced, process entities are relocated, all of these have potential to degrade the performance of the process. Research here could proceed along three fronts. First, the development of heuristics for when such changes should trigger a process revision, or at least an analysis. Second, the design of processes that are robust to process changes. Third, the development of practical measures of process degradation that could inform economic models for process management.

LIMITATIONS

This was an exploratory study and as such can only "paint with broad strokes." Since data was drawn from three distinct regions but combined into one data set, any summary statistics are composite and may not be representative of one specific region. However, the goal of the survey was not to establish a specific percentage of improvement activity that is caused by degradation, but rather to learn if the percentage is substantive: it is enough to merit attention? The study suggests that one fourth to one third of changes made in an improvement project correct degradation and that the corrections being made have a substantive effect on productivity. Improved management of process degradation holds potential benefits..

While the data set at first appears to be large, with data on 351 improvement projects, the subdivisions by sector, years of CI, improvement perspective, change type, degradation type, and so forth quickly lead to small n values for specific pivots. A larger data set would be necessary to make conclusions and to uncover insights that might be found through factor comparisons.

There are some limitations associated with the choice of the improvement *project* as the unit of study. Changes are made in processes that do not occur through the mechanism of a project. For example, (Akers, 2016) reports a practice at FastCap, LLC, whereby individual workers make one small improvement each day. Improvements resulting from operations research, such as optimal schedules, optimal work allocations, optimal replacement timings, are other examples. The reported results are limited to improvement in a project context, even if the discussion of process degradation and its prevention can be applied in a wider context.

The survey listed 16 change types, a large number to include in the survey, but some respondents selected ***other***, demonstrating that the list was not exhaustive. More attention is needed to determine other categories of change to include. For example, the principle, "replace human effort with machine effort (including computers) to increase throughput and quality" only appeared in the list as applied to software, not physical manufacturing processes. Finally, the list may have been biased by the background of the authors. Also, the approach of having the respondents select those changes that occurred in the project, as a yes or no, caused a simplification of the data that prevented a more accurate quantification of the impact of a given change.

Some of the wording in the survey was problematic. Question 5, which asked to estimate the percentage of changes that were ***corrective or standard*** and the percentage of those that were ***creative or novel*** did not allow a direct measure of the percentage of changes that corrected degradation. The question should be improved in future study.

APPENDIX A: Survey Questions

Project Title ______________________________________(for your reference; not used in the study)

For this project

1. What prompted this improvement project? (circle all answers that apply)
   a. quality issues (defects)
   b. safety issues
   c. profitability issues (poor productivity, time wasted, high process costs)
   d. benchmarking with other firms
   e. new technology available
   f. frustrated employees
   g. compliance concerns
   h. innovative thinking about how to be more competitive
   i. management's project quota
   j. other ___________________________

2. Before the improvement project was started, how would you describe the process performance? (circle all that apply)
   a. performance had declined due to wear and tear on equipment.
   b. performance had declined due to product changes that affected the process.
   c. performance had declined because we replaced some tools, machines, software, or workers and things weren't working as well together anymore
   d. growth in demand on the process was pushing it against capacity limits.
   e. growth in demand on the process was causing performance issues.
   f. the process was working as well as ever.

3. Before the improvement project was started, how would you describe the process itself? (circle all that apply)
   a. there were too many different flow paths through the process
   b. the work required of the various people or machines in the process was not well balanced (some people or machines had too much to do, others not enough).
   c. the workplace was disorganized and cluttered, it was confusing to work in.
   d. various components of the process did not seem to work well together
   e. the data stored in our computers or in people's minds did not match reality
   f. we were not able to provide as much customization as customers desired
   g. the throughput time of the process was too long
   h. the process was not in compliance with relevant regulations
   i. none of these were true

4. What was changed to improve the process? (circle all that apply)
   a. we fixed various problems with the equipment
   b. we replaced equipment with basically identical equipment
   c. we replaced equipment with newer equipment, same technology, but better
   d. we changed to a new technology, a different way of processing
   e. we redesigned the product itself, to better the process
   f. we made work assignment changes
   g. we moved things around physically or in sequence of steps to make the flow better
   h. we added capacity
   i. we removed unnecessary equipment, material, and tools from the workplace/process
   j. we reorganized the workplace
   k. we added signage or visual control elements
   l. we put error proofing devices in place (to protect the product)
   m. we put safety devices in place (to protect the employee)
   n. we added fixtures or jigs to make the task more efficient
   o. we added some new features to the software to make the task more efficient
   p. we changed to software with a better (more efficient) user interface
   q. other _________________________

5. Compare the process before the improvement to that after the improvement. Roughly estimate what percentage of the changes made were basically corrective or standard and what percentage were creative or novel.

   ________ % corrective or standard ________ % creative or novel

6. Were any of the following perspectives employed? (circle all that apply)
   a. Six sigma
   b. a kaizen event
   c. lean
   d. 5S
   e. TPM
   f. industrial engineering
   g. benchmarking
   h. TQM
   i. bottleneck or constraint management
   k. other ___________________________

7. Indicate the amount of improvement experienced by this project in each of the various categories. (circle a,b,c, or d in each column)

productivity	**quality**	**on time delivery**	**safety**	**ability to customize**	**competitiveness**
a. a change for the worse	a. a change for the worse	a. a change for the worse	a. a change for the worse	a. a change for the worse	a. a change for the worse
b. no change	b. no change	b. no change	b. no change	b. no change	b. no change
c. a little improvement	c. a little improvement	c. a little improvement	c. a little improvement	c. a little improvement	c. a little improvement
d. significant improvement	d. significant improvement	d. significant improvement	d. significant improvement	d. significant improvement	d. significant improvement

Organization information

I. Our industry is:
- a. manufacturing
- b. financial services (including insurance)
- c. hospitality and retail
- d. healthcare
- e. education and consulting
- f. other services
- g. other ________________________ .

II. Organization size
- a. under 20 employees
- b. 20 to 50 employees
- c. 50 to 100 employees
- d. 100 to 500 employees
- e. 500 to 2000 employees
- f. over 2000 employees

III. Organization type
- a. private
- b. public
- c. not for profit

IV. Our organization has been doing formal continuous improvement projects for
- a. 1 to 2 years
- b. 3 to 5 years
- c. 6 to 10 years
- d. 11 to 20 years
- e. over 20 years

V. Approximately how many CI projects does your location of your organization do per year? _______

VI. How would you characterize the continuous improvement function within your organization?
- a. young and interesting, gaining support
- b. well accepted and healthy
- c. well accepted and ho-hum.
- d. in danger of losing management or employee support
- e. other ____

VII. How would characterize the focus of CI projects in your organization on a scale from solving process problems to developing innovative processes? (circle a plus sign.)

problem solving +----------+----------+----------+----------+----------+----------+----------+ innovating

Note: Each respondent was given three copies of the survey page (questions 1 through 7) and was instructed to fill out the form for three recent improvement projects at the organization.

APPENDIX B – Odds ratio tables

Table B1: Odds Ratios for Significant Variables Predicting Quality Improvement

Type	Change	% of projects	**Odds Ratio**	OR 2.5%	OR 97.5%	p value
main	reconfigured flow - maintenance	14%	**2.2**	1.2	4.2	0.017
main	added capacity	19%	**1.6**	1.0	2.7	0.075
main	decluttered workplace - maintenance	14%	**2.0**	1.1	3.8	0.030
main	other -maintenance	5%	**3.4**	1.4	8.4	0.007
enh	moved to new technology	31%	**2.9**	1.8	4.7	0.000
enh	added fixtures or jigs	19%	**1.9**	1.1	3.4	0.028
enh	added features to software	30%	**2.4**	1.5	4.1	0.001
enh	better user interface	25%	**0.6**	0.4	1.1	0.090

Table B2: Odds Ratios for Significant Variables Predicting On-time Delivery Improvement

Type	Change	% of projects	**Odds Ratio**	OR 2.5%	OR 97.5%	p value
main	fixed equipment	14%	**1.7**	0.9	3.0	0.089
main	changed work assignment - maintenance	35%	**2.1**	1.3	3.3	0.002
main	reconfigured flow - maintenance	14%	**4.0**	2.1	8.0	0.000
main	reorganized the workplace - maintenance	19%	**1.8**	1.0	3.2	0.048
main	other -maintenance	5%	**2.3**	0.9	5.8	0.073
enh	moved to new technology	31%	**2.0**	1.3	3.2	0.002
enh	reconfigured flow - enhancement	23%	**1.7**	1.0	2.9	0.036
enh	added safety measures	13%	**0.6**	0.3	1.1	0.080
enh	other -enhancement	3%	**2.7**	0.9	8.4	0.071

Table B3: Odds Ratios for Significant Variables Predicting Safety Improvement

Type	Change	% of projects	**Odds Ratio**	OR 2.5%	OR 97.5%	p value
main	fixed equipment	14%	**2.7**	1.5	5.2	0.002
main	reorganized the workplace - maintenance	19%	**2.5**	1.4	4.6	0.002
enh	changed work assignments - enhancement	15%	**0.3**	0.1	0.7	0.006
enh	reorganized the workplace - enhancement	21%	**2.3**	1.3	4.2	0.006
enh	added safety measures	13%	**7.7**	3.9	15.5	0.000
enh	added fixtures or jigs	19%	**1.7**	1.0	3.0	0.062
enh	better user interface	25%	**0.5**	0.3	0.9	0.025

Table B4: Odds Ratios for Significant Variables Predicting Customization Improvement

Type	Change	% of projects	**Odds Ratio**	OR 2.5%	OR 97.5%	p value
enh	moved to new technology	31%	**1.5**	1.0	2.3	0.070
enh	redesigned the product	28%	**1.6**	1.1	2.6	0.029
enh	added visual cues	28%	**1.6**	1.0	2.5	0.051
enh	added features to software	30%	**2.0**	1.3	3.2	0.002

Table B5: Odds Ratios for Significant Variables Predicting Competitiveness Improvement

Type	Change	% of projects	Odds Ratio	OR 2.5%	OR 97.5%	p value
main	fixed various problems with the equipment	14%	2.7	1.4	5.0	0.002
main	added capacity	19%	1.9	1.1	3.3	0.015
main	reorganized the workplace - maintenance	19%	3.0	1.7	5.5	0.000
enh	reorganized the workplace - advancement	21%	2.2	1.3	3.7	0.005
enh	better user interface	25%	1.7	1.1	2.9	0.030

REFERENCES

Ahuja, I. P. S. and Khamba, J. S. (2008). Total productive maintenance: literature review and directions. *International Journal of Quality & Reliability Management*, 25(7), August, 709–756.

Akers, P. A. (2016). *2 Second Lean-: How to Grow People and Build a Fun Lean Culture*. BookBaby.

Best, M. and Neuhauser, D. (2006). Walter A Shewhart, 1924, and the Hawthorne factory. *BMJ Quality & Safety*, 15(2), April, 142–143.

Bhuiyan, N. and Baghel, A. (2005). An overview of continuous improvement: from the past to the present. *Management Decision*, 43(5), June, 761–771.

Boer, H., Berger, A., Chapman, R., and Gertsen, F. (2017). *CI changes from suggestion box to organisational learning: continuous improvement in Europe and Australia: Continuous Improvement in Europe and Australia*. Routledge.

Camp, R. C. (1989). *Benchmarking: the search for industry best practices that lead to superior performance*. ASQ Press.

Chakraborty, A. and Chuan, T. K. (2013). An empirical analysis on Six Sigma implementation in service organisations. *International Journal of Lean Six Sigma*, 4(2), May, 141–170.

Deming, W. E. (1986). *Out of the crisis*. Massachusetts Institute of Technology, Center for Advanced Engineering Study.

Desai, D. A., Antony, J. and Patel, M. B. (2012). An assessment of the critical success factors for Six Sigma implementation in Indian industries. *International Journal of productivity and performance management*, 61(4), April, 426–444.

Doolen, T. L. and Hacker, M. E. (2005). A review of lean assessment in organizations: an exploratory study of lean practices by electronics manufacturers. *Journal of Manufacturing systems*, 24(1), January, 55–67.

Emerson, H. P. and Naehring, D. C. E. (1988). *Origins of industrial engineering: the early years of a profession*. Industrial Engineering & Management Press.

Fujimoto, T. (1999). *The evolution of a manufacturing system at Toyota*. Oxford University Press.

Gantt, H. L. (1901). A bonus system of rewarding labor. *Transactions of the American Society of Mechanical Engineers*, 23, December, 341–372.

Goldratt, E. M. and Cox, J. (2004). *The goal: a process of ongoing improvement*. rev, 20th. Great Barrington, MA: North River Press.

Harrington, H. J. (1995). Continuous versus breakthrough improvement: finding the right answer. *Business Process Re-engineering & Management Journal*, 1(3), December, 31–49.

Harry, M. and Schroeder, R. (1999). *SIX SIGMA: The breakthrough management strategy revolutionizing the world's top corporations 1st ed.* Doubleday Business.

Hirano, H. (1995). *5 pillars of the visual workplace*. Productivity Press.

Holweg, M., Davies, J., De Meyer A., Lawson, B. and Schmenner, R.W. (2018). *Process theory: The principles of operations management*. Oxford University Press.

Holweg, M., Staats, B. R. and Upton, D. M. (2018). Making process improvements stick, *Kenan Institute of Private Enterprise Research Paper*, (18–22).

Hosmer, D. W. and Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods*, 9(10), January, 1043–1069.

Huntzinger, J. (2002). The roots of lean. *Training Within Industry: The Origin of Kaizen, Association for Manufacturing Excellence*, 18(2), 14–23.

Imai, M. (1986). *Kaizen*. Random House Business Division.

Keegan, R. (2011). *Becoming Lean: Practical Steps to Build Competitiveness*. NuBooks.

Khamalah, J. N. and Lingaraj, B. P. (2007). TQM in the service sector: a survey of small businesses. *Total quality management*, 18(9), November, 973–982.

Kitaw, D. and Belachew, T. (2007). Work simplification for productivity improvement. A case study: Kaliti Metal Products Factory. *Zede Journal*, 24, 43–57.

Kobayashi, K., Fisher, R. and Gapp, R. (2008). Implementing 5S within a Japanese context: an integrated management system. *Management Decision*, 46(4), May, 565–579.

Kovach, J. V, Cudney, E. A. and Elrod, C. C. (2011). The use of continuous improvement techniques: A survey-based study of current practices. *International Journal of Engineering, Science and Technology*, 3(7), 89–100.

Kumar, M., Antony, J. and Douglas, A. (2009). Does size matter for Six Sigma implementation?

Findings from the survey in UK SMEs. *The TQM journal*, 21(6), October, 623–635.

Laraia, A. C., Moody, P. E. and Hall, R. W. (1999). *The kaizen blitz: accelerating breakthroughs in productivity and performance*. John Wiley & Sons.

Lewis, M. A. (2007). Charles Babbage: Reclaiming an operations management pioneer. *Journal of Operations Management*, 25(2), March, 248–259.

Liker, J. K. (1997). *Becoming lean: inside stories of U.S. manufacturers*. Productivity Press.

Lipsitz, S. R., Fitzmaurice, G. M. and Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 45(2), June, 175–190.

Marcosson, I. F. (1945). *Wherever Men Trade: The Romance of the Cash Register*. Dodd, Mead.

Melnyk, S.A., Calantone, R. J., Montabon, F.L. and Smith, R.T. (1998). Short-term action in pursuit of long-term improvements: Introducing Kaizen events. *Production and Inventory Management Journal*, 39(4), October, 69-76.

Meyer, B. C. and Bishop, D. S. (2007). Florence Nightingale: Nineteenth century apostle of quality. *Journal of Management History*, 13(3), July, 240-254.

Meyer, B., Bumblauskas, D., Keegan, R. and Zhang, D. (2023). When bad things happen to good processes: a theory of entropy for process science. *International Journal of Productivity and Performance Management*, ahead-of-print.

Mogensen, A. H. and Rausa, R. Z. (1989). *Mogy: an autobiography:" father of work simplification"*. Idea Associates.

Nakajima, S. (1988). *Introduction to TPM: total productive maintenance.(Translation)*. Productivity Press, Inc.

Niebel, B. W. (1967). *Motion and time study*. RD Irwin.

Nordin, N., Md Deros, B. and Abd Wahab, D. (2010). A survey on lean manufacturing implementation in Malaysian automotive industry. *International Journal of innovation, management and technology*, 1(4), 374–380.

Ohno, T. (1988). *Workplace Management*. Productivity Press.

Osada, T. (1991). *The 5S's: five keys to a total quality environment*. Quality Resources.

Process Mining and Execution Management Software | Celonis (no date). Available at: https://www.celonis.com/ (Accessed: 23 October 2021).

Robinson, A. G. and Robinson, M. M. (1994). On the tabletop improvement experiments of Japan. *Production and operations management*, 3(3), September, 201–216.

Roser, C. (2016). *Faster, Better, Cheaper in the History of Manufacturing: From the Stone Age to Lean Manufacturing and Beyond*. Productivity Press.

Schonberger, R. J. (2007). Japanese production management: An evolution—With mixed success. *Journal of Operations Management*, 25(2), March, 403–419.

Schroeder, D. M. and Robinson, A. G. (1991). America's Most Successful Export to Japan: Continuous Improvement Programs', *Sloan Management Review*, 32(3), April, 67.

Shewhart, W. A. (1926). Quality control charts 1. *Bell System Technical Journal*, 5(4), October, 593–603.

Shimokawa, K. and Fujimoto, T. (2009). *The Birth of Lean: Conversations with Taiichi Ohno, Eiji Toyoda, and Other Figures who Shaped Toyota Management*. Cambridge: Lean Enterprise Institute.

Shingō, S. and Robinson, A. (1990). *Modern approaches to manufacturing improvement: the Shingo system*. Productivity Press.

Sila, I. and Ebrahimpour, M. (2005). Critical linkages among TQM factors and business results. *International journal of operations & production management*, 25(11), November, 1123–1155.

Singh, J. and Singh, H. (2012). Continuous improvement approach: state-of-art review and future implications. *International Journal of Lean Six Sigma*, 3(2), June, 88–111.

Taylor, F. W. (2004). *Scientific management*. Routledge.

Voss, C. A., Åhlström, P. and Blackmon, K. (1997). Benchmarking and operational performance: some empirical results. *International Journal of Operations & Production Management*, 17(10), October, 1046–1058.

Willmott, P. and McCarthy, D. (2001). *TPM-: A Route to World Class Performance*. Newnes.

Womack, J. P., Jones, D. T., and Roos, D. (1990). *The machine that changed the world: based on the Massachusetts Institute of Technology 5-million dollar 5-year study on the future of the automobile*. Rawson Associates.

Yusuf, Y., Gunasekaran, A. and Dan, G. (2007). Implementation of TQM in China and organisation performance: an empirical investigation. *Total quality management*, 18(5), July, 509–530.

Yusuf, Y. Y. and Adeleye, E. O. (2002). A comparative study of lean and agile manufacturing with a related survey of current practices in the UK. *International journal of production research*, 40(17), January, 4545–4562.

DECISION SCIENCES INSTITUTE

The Impact of Supply Chain Strategy and Entrepreneurial Orientation on Supply Chain Resilience

Prashanth Anekal
Saginaw Valley State University
Email: panekal@svsu.edu

Surender Reddy
Saginaw Valley State University
Email: reddy@svsu.edu

ABSTRACT

This paper explores the impact of Supply Chain Strategy and Entrepreneurial Orientation on Supply Chain Resilience in Family-Owned Small and Medium-sized Enterprises (SME). Supply Chain Strategy is conceptualized in terms of Efficiency focused strategy and Agility focused strategy. Efficiency focus is proposed to have a negative effect on resilience, whereas agility focus is proposed to have a positive effect on resilience. Entrepreneurial Orientation is proposed to positively influence Supply Chain Resilience. Theoretical and managerial implications are discussed.

KEYWORDS: Supply chain strategy, Entrepreneurial orientation, Supply chain resilience

INTRODUCTION

SMEs are the backbone of the American economy or the economy of any country in general. SMEs account for a substantial portion of the wealth generated in the country and employ a large percentage of the working population in the country. According to the World Bank, SMEs represent around 90% of businesses and more than 50% of employment worldwide. In the US, small businesses account for 99.9% of all businesses and nearly half of the working population (US Small Business Administration, 2022). These SMEs that are critical for the success of any economy were affected badly during the Covid-19 pandemic. Most of them struggled to keep their businesses operational during the pandemic. Eventually, the pandemic led to the demise of many of these businesses. This has drawn tremendous attention to the topic of resilience within organizations and more so, the resilience of supply chains. Organizations are now looking at ways to make their supply chains more resilient to any such future disruptions or unforeseen events.

Ever since the Covid-19 pandemic, academics and practitioners have renewed their attention on Supply Chain Risk Management and Supply Chain Resilience. This is understandable, given that supply chains span across the globe and global supply chains are now exposed to a variety of external risk factors. Supply chains are now more vulnerable to disruptions than ever before. Risk management literature has previously talked about two factors - probability or likelihood of a disruptive event and the potential impact of that event. Underestimating either of these two factors could have disastrous implications for the supply chain. This is where Supply Chain Resilience comes to the forefront.

Academic research related to Supply Chain Resilience has tried to answer two fundamental questions – (1) What is supply chain resilience; and (2) How can firms and their supply chains become more resilient? The first question has been addressed in numerous studies. The general conceptualization of supply chain resilience is in terms of the ability to be prepared for unexpected or disruptive events, responding to those events, and bouncing back to the original state or in some cases, moving to a much better state (e.g., Christopher and Peck, 2004; Sheffi and Rice, 2005; Ponomarov and Holcomb, 2009). The second fundamental question has also been addressed in many studies in the past. Most of these studies have explored the impact of operational practices on building a more resilient supply chain. For example, the use of internal inventory buffers is seen as a way to increase resilience (Christopher and Peck, 2004; Zsidisin and Wagner, 2010). Pereira et al. (2014) find that procurement activities contribute significantly to increasing supply chain resilience. At the strategic level, supply chain design is said to have an influence on the extent to which supply chains can respond to disruptions without breaking down (Christopher and Peck, 2004; Blackhurst et al., 2011; Speigler et al., 2012). Strategic sourcing practices have also been discussed as a means to make the supply chain more robust and to reduce downtime when faced with unexpected events (Carvalho et al., 2012; Simangunsong, et al., 2012). Despite all of these, there are many unexplored antecedents related to this question. For instance, there are very few studies that explore the role of the overall supply chain strategy and its impact on the resilience of the supply chain. This is addressed as part of the first research question in this study.

In addition to the supply chain strategy question, we also explore the role of Innovation Orientation in enabling a resilient supply chain. Management literature has explored the impact of strategic orientation and its various components on the performance of the firm. Gatignon and Xuereb (1997) identify customer orientation, competitor orientation and technological orientation as the three components of strategic orientation. According to Baker and Sinkula (2009), Market Orientation and Entrepreneurial Orientation are two of the most commonly discussed strategic orientations in literature. Entrepreneurial orientation focuses more on being innovative and proactive. Thus, previous studies have examined the effect of entrepreneurial orientation on new product development (Gatignon and Xuereb, 1997), identifying and exploiting new markets (Miller, 1983), and firm profitability in general (Baker and Sinkula, 2009; Kohli and Jaworski, 1990; Narver and Slater, 1990). Studies in the past have shown that entrepreneurial orientation can help an organization adapt to changes and grow in the face of uncertainty (Hisrich et al., 2016; Hitt et al., 2001), which is an integral part of resilience. There have been other studies which have examined the effect of individual sub-constructs of entrepreneurial orientation on resilience. Golgeci and Ponomarov (2013) find that firm innovativeness is positively associated with supply chain resilience. Oh and Teo (2006) reported a positive relationship between managerial proactiveness and organizational resilience. However, there is still a gap in terms of understanding the effect of entrepreneurial orientation and its various sub-constructs on supply chain resilience.

The rest of the paper is structured as follows. In the next section of the paper, we present the literature review. The proposed relationships (propositions) are in the next section. This is followed by the methodology and discussion in the last two sections.

LITERATURE REVIEW

Supply Chain Strategy

Management literature has focused extensively on corporate strategy and business strategy. Operations management researchers in particular have studied manufacturing strategy. In comparison, supply chain strategy as a focus area is very recent. In fact, the concept of supply chain management itself is relatively new in business / management literature. According to Giunipero et al. (2008), it was introduced in the 1960s by J.W. Forrester, in his book Industrial Dynamics. It was only in the late 1970s and 1980s that it started catching the attention of academic researchers. As businesses started looking overseas for suppliers and customers, they started realizing the importance of supply chain management. The focus on supply chain strategy has increased ever since organizations realized that competition is no longer between firms, but rather between supply chains. Formulating the appropriate Supply chain strategy and ensuring effective supply chain management is the way to go for firms trying to compete in the global marketplace.

One of the most seminal works in the field of supply chain strategy is that of Fisher (1997), published in the Harvard Business Review. Fisher proposed two supply chain strategies – physically efficient strategy and market responsive strategy. He argued that the supply chain strategy must align with the product characteristics and identified two kinds of product types – functional products and innovative products. Fisher's 2 X 2 matrix indicated that based on product characteristics, a physically efficient strategy is better for functional products which have stable demand, long lifecycles, and low product variety. Whereas a market responsive supply chain strategy is well suited for innovative products which have more demand variability, shorter lifecycles, and higher product variety. Childerhouse et al., (2002) and Aitken et al., (2003) tested Fisher's framework and found support for it.

Lee (2002) expanded on Fisher's framework to include both supply uncertainty and demand uncertainty. Lee proposed four supply chain strategies, two of which were the ones Fisher (1997) proposed. The two new ones were Risk hedging strategy and Agile strategy. The risk hedging strategy focuses on pooling and sharing resources in the supply chain in order to reduce risk to individual entities in case of a disruption. This strategy is proposed to be ideal when demand uncertainty is low, but supply uncertainty is high. The agile strategy on the other hand is proposed to be ideal when both supply uncertainty and demand uncertainty are high. In addition to pooling inventory or capital resources, the agile strategy aims to be flexible and responsive to customer needs.

Other studies have taken Fisher's (1997) framework and interpreted it in terms of lean supply chain (e.g., Womack and Jones, 2003) and agile supply chain (Christopher, 2000; Mason-Jones et al., 2000; Yusuf et al., 2004). Agility is said to help when there are disturbances in the external environment and when the market is more dynamic (Lee, 2004). Agile supply chains are also able to reduce lead times in times of uncertainty by leveraging their supply network and the close relations that they have with their supply chain partners (Christopher, 2000).

Previous research has shown that supply chain strategy directly influences supply chain practices and firm practices, which in turn influence firm performance, both operational and financial (Qi et al., 2009; Huang, et al., 2002). Vickery et al., (2003) argue that an integrative supply chain strategy can help improve customer service and increase financial performance in terms of sales and return on investment. Tarafdar et al. (2017) show that an agile supply chain strategy is positively associated with the ability of the supply chain to respond to changing customer demands in terms of volume and other customer product requirements.

In this study, we explore supply chain strategy in terms of an efficiency focused strategy and an agility focused strategy. We then examine how the choice of strategy can have an impact on supply chain resilience. Efficiency focus refers to a supply chain strategy which aims to achieve the highest efficiencies in the supply chain (Fisher, 1997). Agility focus refers to a supply chain strategy which aims to reduce the risk of disruptions or supply shortages and at the same time, aims to be responsive to customer needs (Lee, 2002).

Entrepreneurial Orientation

Strategy and management literature has spent considerable time exploring what constitutes an entrepreneurial firm and examining the difference between an entrepreneurial firm and a non-entrepreneurial firm (Anderson et al., 2015). Among other things, Entrepreneurship Orientation is said to be the key differentiator between them. Entrepreneurial orientation is seen as the driving force behind an organization's entrepreneurial activities (Covin and Wales, 2012). The origins of Entrepreneurship Orientation can be traced back to strategy literature (Mintzberg, 1978). Miller (1983), Khandwalla (1977), Miller and Friesen (1977) and Covin and Slevin (1989) are credited for the initial conceptualizations of Entrepreneurial Orientation. Miller (1983) defined Entrepreneurial Orientation in terms of innovation, proactiveness and risk-taking. According to Wiklund and Shepherd, (2003, p. 246), Entrepreneurship Orientation "represents the policies and practices that provide a basis for entrepreneurial decisions and actions". Some studies have defined Entrepreneurship Orientation in terms of the collective mindset inclined towards innovation, creativity, change and entrepreneurial activities (Rauch et al, 2004; Ma and Tan, 2006). Lumpkin and Dess (1996) expanded on Miller's conceptualization of the three dimensions of entrepreneurial orientation by adding two more dimensions - autonomy and competitive aggressiveness. They define entrepreneurial orientation in terms of processes, practices, and decision-making activities. They describe entrepreneurial orientation as the "propensity to act autonomously, a willingness to innovate and take risks, and a tendency to be aggressive toward competitors and proactive relatively to marketplace opportunities." Merz and Sauber (1995) define entrepreneurial orientation as "the firm's degree of proactiveness (aggressiveness) in its chosen product-market unit (PMU) and its willingness to innovate and create new offerings" (p. 554). In this study, we stick to Miller's (1983) conceptualization of the three dimensions of entrepreneurial orientation. We define Entrepreneurial orientation as the organization's disposition to adopt policies and practices that emphasize proactive, innovative strategies that contain an element of risk (Morris and Paul, 1987).

Innovativeness is seen as a precursor to experimentation, creativity, and research & development (Lumpkin and Dess, 1996). Innovativeness leads to the creation of new products/services or improvements to existing products/services. Innovativeness is especially important when the firm is facing a hostile external environment which requires it to adapt and reconfigure accordingly (Covin and Slevin, 1989).

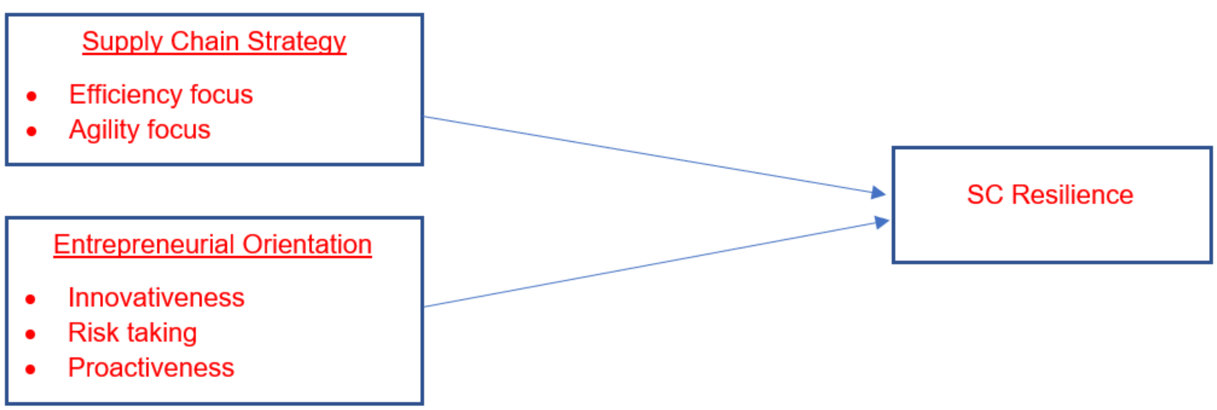
Risk-taking behaviors include taking on debt, accepting risky projects and taking an aggressive position in the face of uncertainty (Miller, 1983). It is said that family businesses generally take on less risk than non-family businesses (Naldi et al., 2007). Family firms tend to be more risk averse and conservative in their approach and strategy making (Naldi et al., 2007; Carney, 2005; Chandler, 1990; Meyer & Zucker, 1989; Schulze et al., 2002). In a study involving family businesses, Naldi et al., (2007) found that risk-taking has a negative impact on firm performance. However, other studies that included risk-taking as a component of entrepreneurial orientation found that firm performance is positively influenced by entrepreneurial orientation (e.g., Li et al., 2008; Wang, 2008; Davis et al., 2010).

Proactiveness is seen as the ability to 'beat your competitors to the punch' and is in essence creating the first mover advantage (Covin and Slevin, 1989; Lumpkin and Dess, 1996). Proactive firms are able to scan the external environment and identify opportunities that give them an upper hand in the marketplace. Proactiveness can also be seen in terms of the firm's ability to adopt new technology, process and practices (Lumpkin and Dess, 1996). Even though proactiveness has been shown to be positively related to firm performance, Lumpkin and Dess (2001) note that the role of proactiveness is heightened under conditions of environmental dynamism. Proactive entrepreneurial behavior is also seen to reap benefits in terms of improved organizational performance (Liem et al., 2019).

Supply Chain Resilience

Supply chain performance has traditionally been judged in terms of cost, quality, delivery reliability, and flexibility. However, in recent times, organizations have been paying a lot more attention to how resilient their supply chains are. Given the risk of disruptions in the external environments, the need to have a more resilient supply chain has increased. Supply chain resilience is now seen as a critical success factor (Liu et al., 2017; Pereira et al., 2014). Previous studies have generally conceptualized supply chain resilience in terms of (1) being prepared to handle disruptions; (2) the ability to bounce back and resume normal operations after a disruption; (3) the ability to return to a much better state after a disruption; and (4) how quickly the supply chain responds to a disruption (e.g., Christopher and Peck, 2004; Rice et al., 2003; Ponomarov et al., 2003; Pereira et al., 2014; Closs and McGarrell, 2004). Shefi and Rice (2005) define a resilient enterprise along the same lines and highlight three key capabilities - readiness, responsiveness, and recovery. Resilient supply chains display both a dynamic capability and an adaptive capability (Gölgeci and Ponomarov, 2013; Kamalahmadi et al., 2016). They must be nimble enough to respond quickly to external changes and be able to reconfigure and evolve when the need arises. Building a resilient system involves an additional cost to the system. The benefits of which may only be realized when faced with a disruption or an interruption in service. Even though disruptions like the one caused by the Covid pandemic may be far and few, disruptions on a smaller scale are much more common. Thus, Supply chain resilience has been an essential component of supply chain literature where researchers have examined it from a risk management perspective by integrating it within the cost-benefit framework (Ponomarov & Holcomb, 2009; Heckman et al., 2015; Juttner and Maklan, 2011).

Figure 1: Research model



HYPOTHESES (PROPOSITIONS) DEVELOPMENT

Supply chain strategy and Supply chain resilience

Supply chain literature has generally talked about two approaches to building resilience – one is by creating operational buffers throughout the supply chain and the other is by creating external supply redundancies (Shekarian et al., 2020; Kamalahmadi et al., 2021; Azadegan et al., 2021). Operational buffering is focused on the internal operations of the organization where the firm tries to maintain excess inventory and excess capacity that they can fall back on when faced with a disruption or interruption in service (Kovach et al., 2015). Creating supply redundancies is mainly about developing alternate sources of supply, engaging in multiple sourcing, and having flexible sourcing contracts (Tomlin, 2006; Qi and Lee 2015). However, both these approaches are resource intensive and result in an additional cost to the system. The concept of operational buffering also goes against the principles of the efficient supply chain strategy where the focus is on achieving cost efficiencies throughout the system. The lean concept which aligns with the efficient supply chain strategy essentially calls for the use of less resources to achieve efficiencies. This approach also attempts to maintain minimal inventory levels since anything more than that is a cost to the system. Lean views inventory sitting on the shelves as money sitting on the shelves and the goal is to reduce that. On the other hand, the agile supply chain strategy, which according to Lee (2002) is a combination of the responsive strategy and risk-hedging supply chain strategy, guards against supply/demand uncertainty by buffering and pooling both inventory and capacity. The agile supply chain strategy helps an organization (and its supply chain) achieve the necessary safeguards against disruptions without going too overboard in terms of additional costs. Thus, we propose:

Proposition 1 – The agile supply chain strategy is positively associated with supply chain resilience, while the efficient supply chain strategy is negatively associated with supply chain resilience.

Entrepreneurial orientation and Supply chain resilience

Management and entrepreneurship literature has in the past examined the differences between firms with a high strategic posture (also known as entrepreneurial firms) and firms with a low strategic posture (also known as conservative firms) (Colvin and Slevin, 1989). The general outlook has been that when faced with hostile environments, entrepreneurial firms perform better than conservative firms (Khandwalla, 1977; Colvin and Slevin, 1989). According to Miller (1983) entrepreneurial efforts are paramount to cope with the adverse forces present in hostile external environments. From an operational perspective, entrepreneurial orientation is said to have a positive effect on the firm's ability to achieve manufacturing flexibility, which is an integral aspect of increasing resilience when faced with the threat of disruptions (Chang et al., 2007). Manufacturing flexibility may be seen in terms of both volume and product mix flexibility. Innovativeness, which is one of the components of Entrepreneurial orientation, plays a key role when firms face disruptions of any sort where the need of the hour may be to reconfigure resources, change product offerings or adopt new processes (Eisenhardt and Martin, 2000; Hult et al., 2004; Jaworski & Kohli, 1993). Gölgeci and Ponomarov (2013) argue that innovative firms are better at responding and adapting to disruptions in the external environment. Innovativeness enables firms to think out of the box, which is critical when facing unexpected situations. Risk-taking, which is the second component of Entrepreneurial orientation, then becomes essential to act upon any innovative solutions that firms may decide on when facing uncertainty. The willingness to take risk then paves that path to devote resources to projects which have uncertain returns. According to Chang et al. (2007), risk-taking practices enable firms to introduce new products, reconfigure existing products, scale up/down production levels and respond to changes in the marketplace. Even though there have been some studies that show that risk-taking is negatively associated with performance, other studies have shown that risk-taking firms perform better during turbulent times than conservative firms (Naldi et al., 2007). Proactiveness, which is the third component of Entrepreneurial orientation, is a forward-looking perspective characteristic of market leading firms which anticipate changes and act ahead of time (Lumpkin and Dess, 2001). Proactive firms seek new opportunities and introduce new products in the market (Miller and Friesen, 1978). Proactiveness is also described as a dynamic capability that enables firms to rebuild competencies in a changing environment (Teece et al., 1997). Previous studies have often found that proactiveness is positively related to firm performance (e.g., Miller 1983; Miller and Friesen 1983). Coleman and Adim (2019) have shown that proactiveness enhances organizational resilience. To summarize, we argue that the three components of entrepreneurial orientation help in making the supply chain more resilient. Thus, we propose:

Proposition 2 – Entrepreneurial orientation is positively associated with supply chain resilience.

METHODOLOGY

This study will test the proposed relationships using survey data collected from SME family-owned manufacturing firms operating in the US. Entrepreneurial orientation will be measured using scales adapted from Covin and Slevin (1989) where innovativeness, risk-taking and proactiveness are measured using two items each. The measurement scale for Supply chain resilience is adapted from Ponomarov (2012) and Golgeci and Ponomarov (2015). The measurement scale for Supply chain strategy will be developed based on previous literature. AMOS package for Structural Equation Modeling (SEM) will be used for data analysis.

DISCUSSION AND CONCLUSION

There is a general sense that supply chains will face increasingly more turbulence in the external environment as time goes on. If that is the case, then we can expect to see more disruptions to global supply chains going forward. This means firms will seek out ways to make their supply chains more resilient to future disruptive events. The next obvious question is how firms can go about doing that? This study thus builds on previous work related to supply chain resilience by identifying two strategic drivers or enablers of supply chain resilience. We expect the results of this study to highlight the role of supply chain strategy and entrepreneurial orientation in the effort toward building supply chain resilience. We argue that the choice of strategy (efficient vs agile) will have an influence on supply chain structure and operational practices that will ultimately help supply chain respond better during disruptions. We also argue that entrepreneurial orientation, which is a form of strategic orientation, will influence the course of action that firms take in their quest to make their supply chains more resilient. In terms of contribution to theory, we build on previous studies that explore the antecedents of supply chain resilience. We will also be developing and testing the scales to measure supply chain strategy, particularly an efficiency focused strategy and an agility focused strategy. In terms of managerial implications, the study intends to throw more light on the effect of strategic choices on resilience and provide more insights to managers who are keen on making their supply chains more effective when faced with disruptions. One of the things we wish to point out as part of this study is that firms need to adopt a balanced approach and not be overly obsessed about efficiency. Firms today are faced with increasingly more cost pressure which results in them seeking out ways to make their systems as lean as possible. However, that may not be the best option when faced with turbulent external environments. A little bit of buffering and redundancy in the supply chain may be essential when seeking a more resilient supply chain. The study also intends to highlight the importance of agility in the long-term survival of the supply chain.

REFERENCES

Baker, W.E. & Sinkula, J.M. (2009). The complementary effects of market orientation and entrepreneurial orientation on profitability in small businesses. Journal of Small Business Management 47: 443–464. DOI:10.1111/j.1540-627X.2009.00278.x

Blackhurst, J., Dunn, K.S. & Craighead, C.W. (2011), "An empirically derived framework of global supply resiliency", Journal of Business Logistics, Vol. 32 No. 4, pp. 374-391.

Carney, M. (2005). Corporate governance and competitive advantage in family-controlled firms. Entrepreneurship Theory & Practice, 29(3), 249–265.

Carvalho, H., Azevedo, S.G. & Cruz-Machado, V. (2012a), "Agile and resilient approaches to supply chain management: influence on performance and competitiveness", Logistics Research, Vol. 4 Nos 1/2, pp. 49-62.

Chandler, A. D. (1990). Scale and scope: The dynamics of industrial capitalism. Boston, MA: Harvard University Press

Christopher, M. (2000). The agile supply chain: Competing in volatile markets. Industrial Marketing Management, 29(1), 37–44.

Christopher, M. & Peck, H. (2004), "Building the resilient supply chain", International Journal of Logistics Management, Vol. 15 No. 2, pp. 1-14.

Davis, J. L., Greg Bell, R., Tyge Payne, G., & Kreiser, P. M. (2010). Entrepreneurial orientation and firm performance: The moderating role of managerial power. American Journal of Business, 25(2), 41-54.

Fisher, M. L. (1997). What is the right supply chain for your product? Harvard Business Review, 75(2), 105–116.

Gatignon, H., & Xuereb, J. M. (1997). Strategic orientation of the firm and new product performance. Journal of marketing research, 34(1), 77-90.

Giunipero, L. C., Hooker, R. E., Joseph-Matthews, S. A. C. H. A., Yoon, T. E., & Brudvig, S. (2008). A decade of SCM literature: past, present and future implications. *Journal of supply chain management*, *44*(4), 66-86.

Hisrich, R. D., Peters, M. P., & Shepherd, D. (2016). Entrepreneurship (10th ed.). McGraw-Hill Education.

Hitt, M. A., Ireland, R. D., Camp, S. M., & Sexton, D. L. (2001). Strategic entrepreneurship: Entrepreneurial strategies for wealth creation. Strategic Management Journal, 22(6–7), 479–491. https://doi.org/10.1002/smj.196

Huang, Z., Li, S. X., & Mahajan, V. (2002). An analysis of manufacturer-retailer supply chain coordination in cooperative advertising. *Decision sciences*, *33*(3), 469-494.

Kohli, A.K. & Jaworski, B.J. (1990). Market orientation: the construct, research propositions, and managerial implications. Journal of Marketing 54: 1–18. DOI:10.2307/1251854.

Lee, H. (2004). The triple-A supply chain. Harvard Business Review, 82(10), 102–112.

Lee, H. L. (2002). Aligning supply chain strategies with product uncertainties. *California management review*, *44*(3), 105-119.

Li, Y. H., Huang, J. W., & Tsai, M. T. (2009). Entrepreneurial orientation and firm performance: The role of knowledge creation process. *Industrial marketing management*, *38*(4), 440-449.

Liem, N. T., Khuong, N. V., & Khanh, T. H. T. (2019). Firm constraints on the link between proactive innovation, open innovation and firm performance. *Journal of Open Innovation: Technology, Market, and Complexity*, *5*(4), 88.

Lumpkin, G. T., & Dess, G. G. (2001). Linking two dimensions of entrepreneurial orientation to firm performance: The moderating role of environment and industry life cycle. *Journal of business venturing*, *16*(5), 429-451.

Mason-Jones, R., Naylor, B., & Towill, D. R. (2000). Lean, agile or leagile? Matching your supply chain to the marketplace. International Journal of Production Research, 38(17), 4061–4070

Miller D. 1983. The correlates of entrepreneurship in three types of firms. Management Science 29: 770–791. DOI:10.1287/mnsc.29.7.770

Miller, D., & Friesen, P. (1978). Archetypes of strategy formulation. Management Science 24:921–933

Meyer, M.W. & Zucker, L.G. (1989).Permanently failing organizations. Newbury Park, CA: Sage Publications.

Naldi, L., Nordqvist, M., Sjöberg, K., & Wiklund, J. (2007). Entrepreneurial orientation, risk taking, and performance in family firms. *Family business review*, *20*(1), 33-47.

Narver JC, Slater SF. (1990). The effect of a market orientation on business profitability. Journal of Marketing 54: 20–35. DOI:10.2307/ 1251757.

Oh, L. B., & Teo, H. H. (2006). The impacts of information technology and managerial proactiveness in building net-enabled organizational resilience. In The Transfer and Diffusion of Information Technology for Organizational Resilience: IFIP TC8 WG 8.6 International Working Conference, June 7–10, 2006, Galway, Ireland (pp. 33-50). Springer US.

Ponomarov, S.Y. and Holcomb, M.C. (2009), “Understanding the concept of supply chain resilience”, International Journal of Logistics Management, Vol. 20 No. 1, pp. 124-143.

Qi, Y., Boyer, K. K., & Zhao, X. (2009). Supply chain strategy, product characteristics, and performance impact: evidence from Chinese manufacturers. *Decision Sciences*, *40*(4), 667-695.

Roberta Pereira, C., Christopher, M., & Lago Da Silva, A. (2014). Achieving supply chain resilience: the role of procurement. Supply Chain Management: an international journal, 19(5/6), 626-642.

Sheffi, Y. & Rice, J.B. Jr. (2005), “A supply chain view of the resilient enterprise”, MIT Sloan Management Review, Vol. 47 No. 1, pp. 41-48.

Simangunsong, E., Hendry, L.C. & Stevenson, M. (2012), “Supply-chain uncertainty: a review and theoretical foundation for future research”, International Journal of Production Research, Vol. 50 No. 16, pp. 4493-4523.

Spiegler, V.L.M., Naim, M.M. & Wikner, J. (2012), “A control engineering approach to the assessment of supply chain resilience”, International Journal of Production Research, Vol. 50 No. 21, pp. 6162-6187.

Vickery, S. K., Jayaram, J., Droge, C., & Calantone, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: an analysis of direct versus indirect relationships. *Journal of operations management*, *21*(5), 523-539.

Vickery, S. K., Jayaram, J., Droge, C., & Calantone, R. (2003). The effects of an integrative supply chain strategy on customer service and financial performance: an analysis of direct versus indirect relationships. *Journal of operations management*, *21*(5), 523-539.

Wang, C. L. (2008). Entrepreneurial orientation, learning orientation, and firm performance. Entrepreneurship theory and practice, 32(4), 635-657.

Womack, J., & Jones, D. (2003). Lean thinking. London: Simon & Schuster

Yusuf, Y. Y., Gunasekaran, A., Adeleye, E. O., & Sivayoganathan, K. (2004). Agile supply chain capabilities: Determinants of competitive objectives. European Journal of Operational Research, 159(2), 379–392

Zsidisin, G.A. & Wagner, S.M. (2010), “Do perceptions become reality? The moderating role of supply chain resiliency on disruption occurrence”, Journal of Business Logistics, Vol. 31 No. 2, pp. 1-20.

DECISION SCIENCES INSTITUTE

The Importance of Project Management Ethics and Its Effects on Business Success

Maling Ebrahimpour
University of Rhode Island
mebrahimpour@uri.edu

Rachel Vennel
University of Rhode Island
rachel_vennel@uri.edu

Chris Orr
University of Rhode Island
orrc215@uri.edu

Paulina Loredo-Vilchis
University of Rhode Island
ploredo19@uri.edu

ABSTRACT

This study explores how ethics impact project management's role in a company's success, focusing on industries like construction, law, software engineering, and the military. Project managers face unique pressures, often making ethically questionable decisions due to time, budget, and scope concerns. However, such choices, like overcharging clients to offset team inefficiencies, have enduring negative impacts. Maintaining an ethical framework leads to improved client relationships and enhanced profitability. Our research suggests integrating ethics training into project management education and emphasizing ethics certification through organizations like PMI.

Key Words: Project Management, Ethics, Certification

DECISION SCIENCES INSTITUTE

The Influence of Power Dynamics on E-mail Communication

Younsoo (Joy), Kim
Columbia University in the City of New York
E-mail: yk2949@columbia.edu

ABSTRACT

This is the first paper that empirically examined the impact of power in actual workplace e-mail, the Enron corpus, using natural language processing (NLP) methodology. The organizational hierarchy between a sender and a receiver is defined as power dynamics in this paper. According to the result, high-power senders tend to write shorter messages to low-power receivers than low-power senders do. Additionally, high-power receivers reply later and are more likely not to respond (left-on-read) to low-power senders than in the opposite communication direction. From the result, this study suggests how to infer latent power dynamics in interpersonal relationships using NLP methods.

KEYWORDS: organizational behavior, natural language processing, computer-mediated communication, explication of relationships, social network theory, cognition and reasoning, and human judgment

INTRODUCTION

Message 1: We should be getting gas fax

Message 2: John, I am planning to come to Houston Wednesday and Thursday to sit with the East desk. Unless that is going to create problems, would you mind letting them know that I'm coming. Chris

Among those two messages from the 500,000 Enron e-mail corpus, which one comes from a boss, and which one goes to a boss? As you already guessed, the first message is from the highest-ranked CEO to the lowest-ranked employee in the hierarchical organizational structure. The second message is opposite from the lowest-ranked employee to the highest-ranked CEO.

According to the Nielson survey, Americans use e-mail for a third of their online communication. Just as in everyday conversation, e-mail is also used in the workplace. In the workplace, power dynamics in a relationship exist. For example, if a manager works for the CEO, the manager is a low-power person(subordinate position), whereas the CEO is a high-power person (dominant position). However, if an employee works for the manager, the same manager is in a high-power position. As you can see, the power dynamic is a context-dependent concept. Because of its abstract characteristics, it is hard to capture the power dynamics and their influence.

The influence of power dynamics has been studied in multiple domains. In psychology, researchers found that power can influence inhibition and approach, expression of negative and positive emotion (Keltner, Gruenfeld, & Anderson, 2003; Hecht and LaFrance, 1998), expression of one's opinion in controversial topics (Berdahl and Martorana, 2006), perspective taking (Galinsky, Magee, Inesi, & Gruenfeld, 2006), creativity (Galinsky, Magee, Gruenfeld, Whitson, & Liljenquist, 2008), and non-conforming behavior (Bellezza, Gino, and Keinan 2014). In linguistics and human-computer interaction, from the early Carasik and Grantham's 1988 paper, researchers found that high-power people use more directives (Vine, 2009) and humor (Holmes and Schnurr, 2005). Most recent research also showed that power can influence commonly used word and phrase choices (Gilbert, 2012), formality(Peterson, Hohensee, & Xia, 2011), and gossip (Mitra and Gilbert, 2013). Organizational hierarchy is also used in social network analysis to detect the power dynamics of a company (Rowe, Creamer, Hershkop, & Stolfo, 2007). In addition to an experimental method to study power dynamics, this paper uses an interdisciplinary approach to measure the influence of hierarchical power in messages with real-world datasets: I used the NLP method to investigate the influence of explicit power on length, left-on-read occurrence, response time, and sentiment with massive Enron e-mail corpus.

I hope that this research lays the groundwork for status-aware computer-mediated communication applications. In organizational behavior, this paper will be beneficial for inferring leaders in a group via mundane discourses. Leaders emerge based on their explicit and/or implicit power dynamics. Knowing who is in a high-power position is essential because this person will be the key player in making a change. Usually, it takes much time to figure out the power dynamics due to their abstract and context-dependent characteristics. However, according to current research, messages and texting between people are enough to infer dominant (high-power) and subordinate (low-power) positions with less cost and time. Also, this research can extend to individual relationships. For example, in a marriage counseling session, text or e-mail messages between a couple can be used to assume who is in the dominant and subordinate positions. Knowing the dominant person in a relationship via message characteristics helps to find a more efficient and actionable solution for a couple in a short time.

LITERATURE REVIEW

Previous research showed that high-power individuals were more likely to express their opinion about a controversial issue during a group discussion than low-power individuals(Berdahl and Martorana, 2006). Bass and Stogdill's 1990 noted that speaking time correlated with leadership: a high-power member could control the speaking time of others. Speaking time shows a high power's dominance and people infer one's high power and status based on speaking time (Mast, 2002; MacLaren et al., 2020). Previous studies used speaking time to infer one's power, but this is the first paper to use the length of written text messages as a signal for one's power. Unlike speaking time, I found that message length negatively correlates with the senders' power.

Left-on-read is a new term that first came out from a mobile instant messaging app. A texting application lets you know when the message is read. If the receiver does not respond even though they read the message, this phenomenon is called "left-on-read." In a romantic relationship, left-on-read can indicate the power dynamics of the low-power sender and high-power receiver. Similar to the romantic relationship, I also found that the left-on-read can occur more frequently between low-power senders and high-power receivers (e.g., an employee sent a mail to a boss) than between high-power senders and low-power receivers (e.g., a boss sent a mail to an employee).

Some previous literature shows the relationship between power and emotional expressions. For example, Van Kleef and Lange 2020 analyzed the effect of power and status on emotional expression. According to their review, high-power people express their emotions more frequently than low-rank people but low-rank people more accurately perceive and respond to emotional expressions. The most similar research paper to my research would be Livingston et al. 2022, which used self-reported narratives to see the relationship between self-reported power, communication about work and sex, and negative emotional expression. They found that individuals with high perceived power tend to talk more about work and show more negative emotions in their narratives. However, this research used the e-mail conversation between actual high versus low-power individuals within the Enron company and analyzed the sentiment of their messages. Unlike the previous findings with experiments and narratives, I do not find statistically significant differences in high-power and low-power individuals' sentiment scores. In addition, I found that only e-mail conversations among the same rank people have lower sentiment than upward or downward communication in the workplace e-mail context.

Creamer et al. 2022 used the response time between the original message and replied to calculate the hierarchy with social network analysis(SNA). However, the paper used the response time as one of the automatic hierarchy detection algorithm components and limited their dataset when the response time was within three business days. On the other hand, I did not limit the response time window to three days. Also, the previous paper used the response time to infer social hierarchy. However, I handled the response time as a dependent variable, which shows the impact of power dynamics between two individuals (one-to-one sender and receiver). Due to the small sample size, the Kruskal-Wallis test was not statistically significant, but the downward reply (reply from boss to employee) tended to take more time than the upward reply(reply from employee to boss).

Even though the previous papers showed the influence of power in experimental methods, I used NLP methodology to analyze the influence of power in actual written communication with the Enron corpus. With the massive dataset, I observed that the dominance showed different patterns in written communication, which showed the opposite direction in previous research papers about the influence of power in speaking: high-power individuals tend to talk more but write less. Moreover, this NLP approach is very parsimonious in capturing the power dynamics between individuals compared to experiments and social network analysis.

METHOD

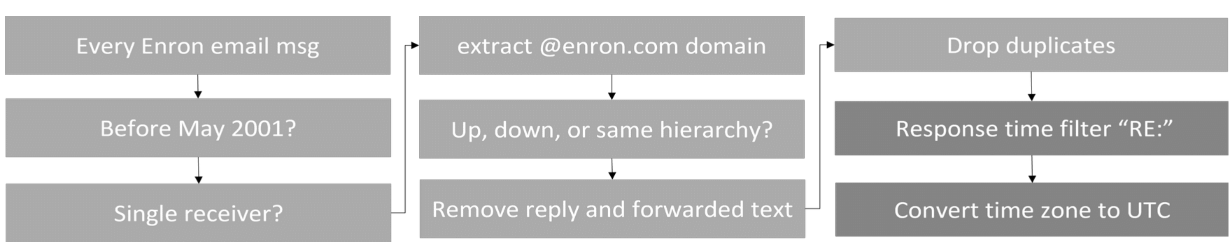


Figure 1: A Map of the steps taken in this paper to prepare the Enron e-mail corpus for analysis

The Enron e-mail corpus(Kliimt and Yang, 2004) and the Enron job title dataset (Peterson, Hohensee, & Xia, 2011) are used to prove the relationship between power dynamics and message characteristics. The Enron e-mail corpus is the largest dataset publicly available to researchers. The dataset contains about 500 thousand e-mails from 151 people over nearly four years. According to Peterson et al., the CEO is in the highest rank (6), whereas the employee, in-house lawyer, and trader are ranked as zero, the lowest.

Figure 1 shows the preprocessing steps. First, following the method of Gilbert's 2012, I discarded the e-mail data after May 1, 2001. According to the history of Enron (Jelveh & Russell, 2006), SEC launched the investigation in October. However, regarding the executives' private selling of Enron stock, May 1 was critical because this was when the executives admitted to themselves about Enron's collapse. This unusual situation was discarded since the research is about normal conversation among the organizational hierarchy.

The main variable of this research is power dynamics between two people, one sender and one receiver. Power dynamics is a relative term. For example, suppose a director sends a message to an employee. In that case, this is a downward communication direction, whereas the same director sending an e-mail to a vice president is an upward communication direction. Even though the sender's rank is the same (director), the receiver's rank changes the communication direction. Therefore, one-to-many (one sender and multiple receivers) is not the focus of the research because it is hard to define the direction. For example, suppose a director e-mailed a vice president and employees simultaneously. In that case, it is hard to decide the direction of communication, whether upward or downward.

The previous research about Enron's organizational hierarchy used the e-mail the company employees exchanged. The social hierarchy information was possible to obtain within the firm based on the employee's job title. However, researchers do not have information outside the company regarding their rank compared to the sender and receiver inside the firm. Therefore, based on the availability of information, I only included the sender and receiver e-mail addresses, whose domains are @enron.com.

With the ranks of sender and receiver, I defined the communication as upward, same, and downward as previous research did (Peterson, Hohensee, & Xia, 2011; Gilbert, 2012). When a receiver outranks the sender, the communication direction is upward, whereas the opposite is recorded as downward. When the ranks of senders and receivers are the same, this communication direction is coded as same. For instance, Hunter Shively was one of the company's vice presidents (rank 4). It was an upward communication direction when he e-mailed other vice presidents (e.g., Andrew Lewis, rank 4). When he sends an e-mail to an employee (rank 0), it is coded downward. When he e-mailed Jeffrey Shankman (president, rank 5), it was an upward communication.

By searching for conventional textual marks (e.g., ---Original Message, ---Forwarded by), I discarded any text that seemed like a quoted reply or a forwarded message. This paper's analysis unit is a digital conversation between two people within a social hierarchy. The original message can be a confounder because it can make the text length longer than it was. To measure the direct communication between sender and receiver, the message after the aforementioned textual mark was deleted.

After filtering, 517,401 observations decreased to 7,131 (with duplicates) and 2021 (without duplicates). I also dropped the duplicates based on sender, receiver, and subject line. In case of response time, I added one more filter: the subject starts with "RE: "or "re:. " To extract the information about when the original and reply messages were sent, I filtered e-mails whose subject start with "RE:" or "re:" Then, to calculate the time difference between the original message and the reply, the recorded time for each message's sent time was converted to UTC (Universal Time Coordinated). The reply messages were recorded in PST(Pacific Standard Time) or PDT(Pacific Daylight Time) zone. For instance, Arnold e-mailed Andy on "Thursday, February 15, 2001, 12:43 PM," and it was replied to "Thu, February 15 2001 05:39:00 -0800 (PST)". To calculate the time difference, the time recorded in PST or PDT was converted to UTC, and the date formats were changed to YYYY-MM-DD HH:MM (e.g, 2001-02-15 13:39, 2001-02-15 12:43). After removing the observation which has a missing value, the total number of e-mails used for response time analysis is 34 messages. I followed the standard procedure for tokenization and removing the stopwords.

RESULTS

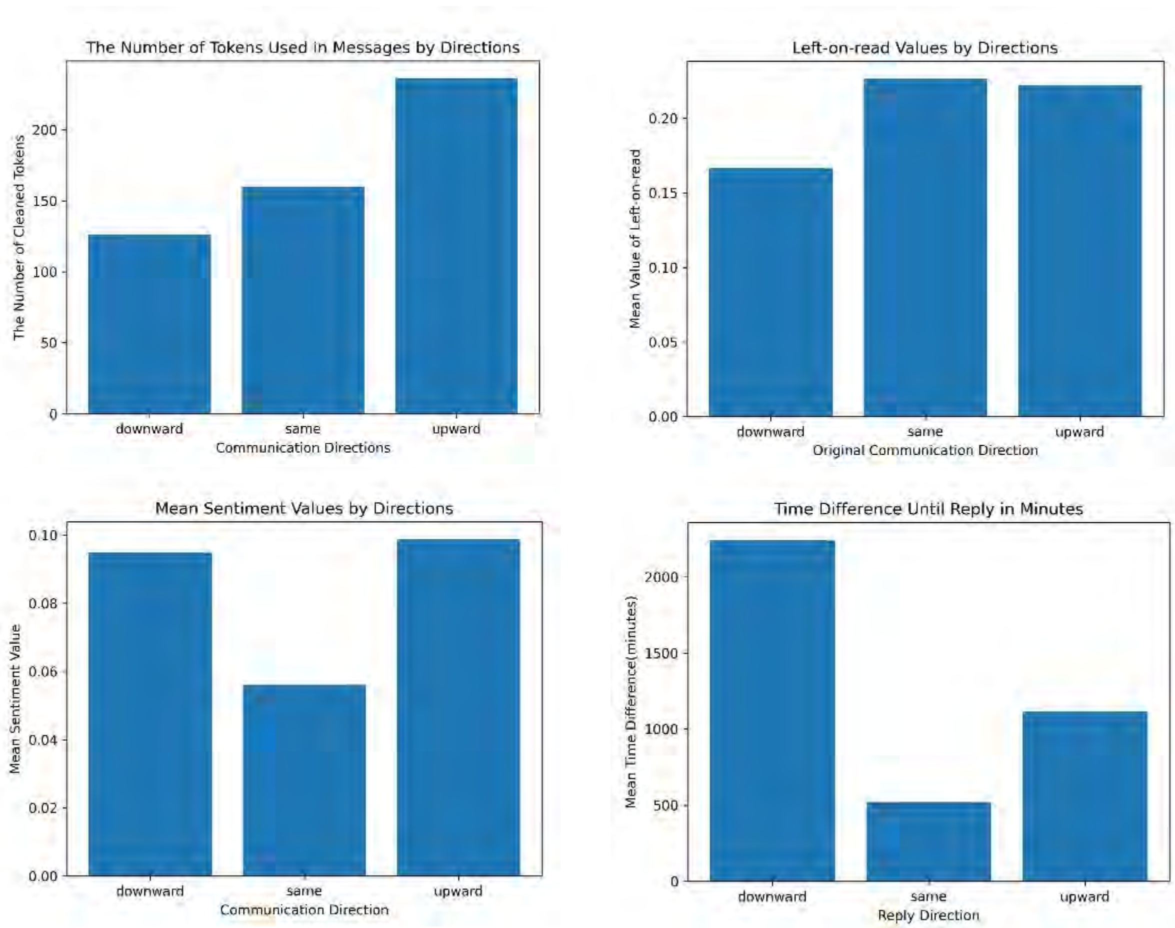


Figure 2: Results of message length, left-on-read, response time, and sentiment

1. Message Length

Each message length between a sender and a receiver was tokenized, and stopwords were removed to quantify the message length based on the number of tokens. One-way ANOVA was conducted to test the hypothesis that there is a significant difference in the message length, represented by the number of tokens of the three communication direction groups. The results of the test indicate a significant overall effect. ($F(2,2020) = 10.99$, $p<0.001$).

The mean number of tokens downward was 126.31(SD=296.34), same was 159.96 (SD=712.34), and upward was 236.50 (SD=708.43). Post-hoc comparison using the Tukey HSD test revealed a significant difference between downward and upward (p = 0.003, C.I. = [29.71, 190.66]) but no significant differences between any other pairwise comparisons.

These results suggest a significant difference in the three communication groups' mean number of tokens used in e-mail communication. For example, when an employee writes an e-mail to the person who outranks the sender (upward communication direction), the message tends to have a longer length than the same (sender and receivers are in the same rank) and downward (sender outranks the receiver).

2. Left-on-read

According to dictionary.com, "left-on-read" is an internet slang that means "when a recipient has read, but not responded to, a sender's message." The authors operationalized the concept of left-on-read as follows: when the number of e-mails sent by authors exceeds the number of reply letters from the recipient, this unbalanced communication is defined as left-on-read. Since I filtered the dataset with the company's domain (@enron.com), I assumed all the e-mails sent to the receiver were read.

For instance, Jeff Dasovich (subordinate low-power; jeff.dasovich@enron.com) was Enron's managing director. He e-mailed vice president James Steffes (superior high-power; james.steffes@enron.com) 61 times. However, the reply from James to Jeff was sent 26 times. This imbalanced communication response is coded as left-on-read.

A one-way ANOVA was performed to compare the effect of communication direction on left-on-read occurrence. As [Figure 3] shows, the one-way ANOVA revealed that there was a statistically significant difference in mean left-on-read occurrent at least two groups ($F(2, 1861) = 4.301$, $p = 0.013$)

Tukey's HSD Test for multiple comparisons found that the mean value of left-on-read occurrence was significantly different between downward and same (p = 0.024, C.I. = [0.006, 0.114]) and downward and upward (p = 0.039, C.I. = [0.002, 0.108]). However, there was no statistically significant difference in the mean left-on-read occurrence between upward and same (p = 0.97).

Overall, the result suggests that communication direction could impact the number of left-on-read, with a downward communication direction resulting in more left-on-read than the same and upward communication direction. So, for example, when an employee sends an e-mail to the boss, it's more likely to be left on read than the boss's e-mail to an employee.

3. Sentiment

The mean sentiment values of downward, same, and upward communication directions were statistically different according to the result of the ANOVA test ($F(2,2020) = 10.98$, $p < 0.001$). In addition, Tukey posthoc test revealed that downward communication was significantly different from the same (p-value <0.001, C.I. = [-0.062, -0.014]). Similarly, the same direction significantly differed from upward communication (p-value <0.001, C.I. = [0.018, 0.066]. However, the downward and upward showed no difference in the mean sentiment score (p-value = 0.925).

The mean value of sentiment in the upward direction is 0.1 (SD = 0.2), in the same was 0.06 (SD = 0.16), and in the downward is 0.09 (SD = 0.2). All of the workplace e-mails had positive sentiments regardless of directions. However, the same group has a slightly negative sentiment compared to other groups.

4. Time delta between original messages and replies

The author extracted the e-mail whose subject starts with "RE: "or "re: "to calculate the time difference between the original messages sent and replied to. Among the dataset, 34 e-mail messages were extracted. The time when the reply was sent was recorded in Pacific Standard Time (PST) or Pacific Daylight saving time (PDT). For comparison, those recorded times were converted to GMT (Greenwich Mean Time, also known as Coordinated Universal Time (UTC+0)).

The mean value of the downward direction was 2244 (SD=3049) minutes, the same was 520 (SD = 486) minutes, and the upward was 117 minutes (SD=419). Based on the mean value, high-power receivers tend to reply lately to low-power senders (2244 minutes). On the other hand, low-power receivers tend to reply quickly to high-power senders (1117 minutes). The response time between the same direction has the smallest value.

Due to the limited number of samples in each communication direction, a Kruskal-Wallis test was conducted to compare the median time difference between the original message sent and the reply. Even though the mean value showed some tendency, the test failed to reveal a significant difference in the time difference between the original and reply message among the three communication directions (H =2.52, df=2, p=0.28)

DISCUSSION AND CONCLUSION

In this paper, I investigated the influence of power dynamics on the e-mail contents of the Enron company. With the Enron corpus and status dataset, which has stratified rank and hierarchy, I can divide the communication direction into three: upward (low-power sender, high-power receiver), same (same rank sender and receiver), and downward (high-power sender and low-power receiver). With the Enron datasets, I figured out that power dynamics can influence message length, left-on-read, response time, and sentiment. When the communication direction is downward (high-power sender and low-power receiver), the sender tends to write short messages and a low likelihood of left-on-read, which are statistically significant. Sentiment does not show a statistically significant difference between upward and downward communication, but the same group has the lowest sentiment, which is statistically significant. For the response time, the data showed a tendency for longer response time in the originally upward direction, but due to the limitation of the sample, the result was not statistically significant.

This research used massive real world e-mail conversation from the Enron corpus. Due to the dataset's characteristics, this study showed a correlational result. Previous research (Smith and Trope, 2006; Wigboldus and A.P Dijksterhuis, 2008) showed that power could shape people's style of thought. Those papers about construal level theory showed that high power tends to make people think abstractly, whereas low power makes people think concretely. Abstract thinking will make people write shorter messages than concrete thinking. So, in the future study, I will use the concreteness dictionary (Brysbaert, Warriner, & Kuperman, 2014) to see if power actually makes the messages more abstract and if the thinking style mediates the message length.

Gossip is one of the possible explanations for the low sentiment score in the same communication direction. Martinescu et al. 2019 showed that hierarchical power could shape individuals' gossip motives and behavior. According to their paper, gossip is less likely to occur in downward than upward and lateral interactions. As most of the gossip has negative sentiment, this can explain why the sentiment score is low only for the same communication direction. As a future study, I will conduct additional topic modeling for whether the contents of e-mail communication are about a personal or work-related.

To understand the influence of power in an e-mail, I asked ChatGPT to generate an example e-mail with the same content (e.g., stock inventory checking) from a boss to an employee and vice versa. The letter to the boss used 270 cleaned tokens, whereas the letter to the employee used 233 cleaned tokens. According to my study, the number of tokens significantly differs among upward and downward directions; the upward direction used nearly twice as many tokens as those in the downward direction. However, ChatGPT did not have much difference in the number of tokens in upward and downward communication directions. It seems the NLP program does not fully consider the influence of power dynamics in written communication. Therefore, power will be a new moderator for the NLP generator with many research opportunities.

REFERENCES

Bass, B. M., & Stogdill, R. M. (1990). *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications.* Simon and Schuster.

Bellezza, S., Gino, F., & Keinan, A. (2014). The red sneakers effect: Inferring status and competence from signals of nonconformity. *Journal of consumer research,* 41(1), 35-54.

Berdahl, J. L., & Martorana, P. (2006). Effects of power on emotion and expression during a controversial group discussion. *European journal of social psychology*, 36(4), 497-509.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods,* 46, 904-911.

Carasik, R. P., & Grantham, C. E. (1988, May). A case study of CSCW in a dispersed organization. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 61-66).

Cho, M., & Keltner, D. (2020). Power, approach, and inhibition: Empirical advances of a theory. *Current Opinion in Psychology,* 33, 196-200.

Creamer, G. G., Stolfo, S. J., Creamer, M., Hershkop, S., & Rowe, R. (2022). Discovering Organizational Hierarchy through a Corporate Ranking Algorithm: The Enron Case. *Complexity*, 2022.

Galinsky, A. D., Magee, J. C., Gruenfeld, D. H., Whitson, J. A., & Liljenquist, K. A. (2008). Power reduces the press of the situation: implications for creativity, conformity, and dissonance. *Journal of personality and social psychology,* 95(6), 1450.

Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological science*, 17(12), 1068-1074.

Gilbert, E. (2012, February). Phrases that signal workplace hierarchy. *In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1037-1046).

Hecht, M. A., & LaFrance, M. (1998). License or obligation to smile: The effect of power and sex on amount and type of smiling. *Personality and Social Psychology Bulletin*, 24(12), 1332-1342.

Holmes, J., & Schnurr, S. (2005). Politeness, humor and gender in the workplace: Negotiating norms and identifying contestation.

Jelveh, Z., & Russell, K. (2006). Interactive timeline: The rise and fall of enron*. The New York Times.*

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological review*, 110(2), 265.

Livingston, T. N., Vik, T. A., & Singer, J. (2022). Relationships Between Power, Communication About Work and Sex, and Emotion Expression: A Linguistic Inquiry and Word Count Analysis. *Psychological Reports*, 00332941221137243.

MacLaren, N. G., Yammarino, F. J., Dionne, S. D., Sayama, H., Mumford, M. D., Connelly, S., ... & Ruark, G. A. (2020). Testing the babble hypothesis: Speaking time predicts leader emergence in small groups. *The Leadership Quarterly*, 31(5), 101409.

Martinescu, E., Janssen, O., & Nijstad, B. A. (2019). Gossip as a resource: How and why power relationships shape gossip behavior. *Organizational Behavior and Human Decision Processes*, 153, 89-102.

Mast, M. S. (2002). Dominance as expressed and inferred through speaking time: A meta-analysis. *Human Communication Research*, 28(3), 420-450.

Mitra, T., & Gilbert, E. (2013). " Analyzing gossip in workplace e-mail" by Tanushree Mitra and Eric Gilbert, with Ching-man Au Yeung as coordinator. *Acm Sigweb Newsletter*, (Winter), 1-7.

Peterson, K., Hohensee, M., & Xia, F. (2011, June). E-mail formality in the workplace: A case study on the Enron corpus. *In Proceedings of the Workshop on Language in Social Media* (LSM 2011) (pp. 86-95).

Rowe, R., Creamer, G., Hershkop, S., & Stolfo, S. J. (2007, August). Automated social hierarchy detection through e-mail network analysis. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 109-117).

Smith, P. K., & Trope, Y. (2006). You focus on the forest when you're in charge of the trees: power priming and abstract information processing. *Journal of personality and social psychology*, 90(4), 578.

Smith, P. K., Wigboldus, D. H., & Dijksterhuis, A. P. (2008). Abstract thinking increases one's sense of power. *Journal of Experimental Social Psychology*, 44(2), 378-385.

Van Kleef, G. A., & Lange, J. (2020). How hierarchy shapes our emotional lives: Effects of power and status on emotional experience, expression, and responsiveness. *Current opinion in psychology*, 33, 148-153.

Vine, B. (2009). Directives at work: Exploring the contextual complexity of workplace directives. *Journal of Pragmatics*, 41(7), 1395-1405.

Wire, N. (2010). What Americans do online: Social media and games dominate activity. *Nielsen Wire, http://blog. nielsen. com/nielsenwire/online_mobile/what-americans-do-online-social-media-and-games-dominate-activity*.

DECISION SCIENCES INSTITUTE

The Intellectual Structure of Social Engineering Research: A Co-occurrence Approach

Anil Singh
Department of Information Systems & Operations Management
University of Texas at Arlington
E-mail: anil.singh@uta.edu

George Mangalaraj
College of Business and Technology
Western Illinois University
E-mail: G-Mangalaraj@wiu.edu

Aakash Taneja
School of Business
Stockton University
E-mail: aakash.taneja@stockton.edu

ABSTRACT

Recent years have witnessed an exponential increase in cybersecurity-related social engineering attempts in various forms. Hence, an understanding of social engineering research in the cybersecurity area is imperative. To this end, we review extant research under the broad umbrella term social engineering. The research included 819 articles from ISI's Web of Science databases. This study analyzes literature through co-word analysis, a traditional bibliometric technique, to unravel the research's seven overarching themes. Our findings provide valuable insights into research depth and direction concerning social engineering research in cybersecurity.

KEYWORDS: Social Engineering, Phishing, Cybersecurity, Co-word analysis.

INTRODUCTION

In recent years, cybersecurity issues have become paramount to society. Among the cybersecurity threats, social engineering is emerging to be a major threat as it takes advantage of the organization's most valuable resource (Mashtalyar et al., 2021). While the concept of Social engineering has existed within the social sciences domain long before the advent of computers, the ubiquitous role of computers in our daily lives has witnessed the migration of Social engineering into the computing domain (Hatfield, 2018).

In the social sciences domain, the term social engineering is typically used in the context of centralized planning to attempt social change and regulate society. In the context of cybersecurity, social engineering is defined as "*the use of social disguises, cultural ploys, and psychological tricks to get computer users to assist hackers in their illegal intrusion or use of computer systems and networks*" (Abraham & Chengalur-Smith, 2010). Likewise, Aldawood and Skinner (2019) define social engineering as a *"method that seeks to exploit a weakness in human nature and take advantage of the naivety of the average person."* Owing to the ubiquity of information technologies and the Internet, the scope and scale of deceptions are worldwide (Steinmetz et al., 2021).

Social engineering attacks cause more significant harm in many ways. For example, the financial impact of social engineering attacks is estimated to cost $120 billion for US companies (Salahdine & Kaabouch, 2019). Some researchers have even claimed that social engineering costs more than natural disasters (Matyokurehwa et al., 2022). Besides financial impacts, social engineering attacks can lead to the theft of intellectual property and confidential information (Mashtalyar et al., 2021).

Research on social engineering in cybersecurity covers a wide range of topics, including preventing social engineering attacks, detecting social engineering attacks, and educating users about social engineering attacks. Hence, research spans technical, socio-technical, and non-technical areas regarding cybersecurity and social engineering. Moreover, social engineering attacks come in many forms, with phishing being one of the popular modes of such attacks. However, other cybersecurity attacks rely on social engineering concepts such as deception to trigger user actions in ransomware, malware, and advanced persistent threat attacks. With researchers studying various aspects of social engineering for the past three decades, there is a need as well as opportunity to analyze the intellectual structure of social engineering research. To this end, our work addresses the following research question:

Research Question: What are the dominant themes – and their interrelationships – in cybersecurity-related social engineering research?

Addressing this research question will allow us to understand the breadth and depth of research on cybersecurity-related social engineering. Moreover, this study's findings will unravel the major themes of research in cybersecurity-related social engineering research. Finally, this study's findings can help find gaps and opportunities for further research on social engineering.

LITERATURE REVIEW

Social Engineering is a weapon of choice among malicious hackers (Abraham & Chengalur-Smith, 2010; Siadati et al., 2017) and poses a serious risk to Information Systems (Krombholz et al., 2015). While systems may be secure, social engineering uses trickery to obtain access and accomplish their nefarious acts (Salahdine & Kaabouch, 2019). Social engineering attacks are prevalent and impact the organization in various ways. Table 1 captures some of the recent social engineering attacks captured by the news media.

Table 1: Representative list of social engineering attacks in recent history.

DATE	ORGANIZATION	IMPACT	SOURCE
Jan. 2023	MailChimp	Some customer information was compromised	https://techcrunch.com/2023/01/18/mailchimp-hacked/
Jan. 2023	Riot Games	Pauses update to the games	https://www.pcmag.com/news/hackers-hit-riot-games-forcing-it-to-halt-game-updates
Sep. 2022	Rockstar Games	The source code of games was stolen	https://news.sophos.com/en-us/2022/09/22/uber-rockstar-fall-to-social-engineering-attacks-and-you/

Sep. 2022	Uber	A hacker gained access to corporate databases and the bug bounty program account	https://www.theverge.com/2022/9/16/23356959/uber-hack-social-engineering-threats
July 2022	Marriott International	A single hotel data was stolen	https://venturebeat.com/security/marriott-social-engineering/

With the prevalence of cybersecurity-related social engineering attacks, researchers have explored various aspects of it. Hence, there have been literature reviews on specific areas of social engineering in the past. For example, Syafitri et al. (2022) reviewed prior literature on social engineering attacks and focused on the prevention methods, models, and frameworks for it. Likewise, Salahdine and Kaabouch (2019) performed a comprehensive review of social engineering attacks and classified them along with the detection and prevention strategies. Yasin et al. (2019) reviewed various social engineering attack scenarios and the persuasive techniques used by the attacker. With the prevalence of social engineering attacks during the COVID pandemic, researchers have reviewed research during that time frame and found prevalent attack types such as phishing, scamming, spamming, smishing, and vishing along with the method of exploits such as fake emails, websites, and mobile apps (Hijji & Alam, 2021).

Apart from literature reviews on social engineering attacks, researchers have also reviewed the research on specific social engineering attacks such as phishing. For example, Gupta et al. (2016) surveyed the literature on phishing attacks that included types of phishing attacks such as tab-napping, spoofing emails, and Trojan horses, among others, and analyzed them and provided detection techniques. Likewise, a review of phishing detection approaches has also been performed by (Abu Zuraiq & Alkasassbeh, 2019).

As mentioned above, many literature review studies on specific aspects of social engineering attacks exist. However, none of these studies provide a comprehensive overview of the corpus of research on cybersecurity-related social engineering. As such, researchers have highlighted technical and non-technical aspects of cybersecurity practices (Malatji et al., 2020), and social engineering is no exception as there are both technical, non-technical, and socio-technical aspects in it. On the non-technical side of research, for example, Luo et al. (2011) have used criminology and social psychology to discuss the personality traits related to social engineering attacks. Likewise, Siddiqi et al. (2022) studied the psychology of social engineering to explore human vulnerabilities employed by criminals. On the technical side of research on social engineering, Asiri et al. (2023) review current state-of-art deep learning models to detect URL-based and hybrid-based phishing attacks. Edwards et al. (2017) propose an automated social engineering vulnerability scanner that organizations can use. Socio-technical approaches, which use social and technical aspects, have created the most powerful weapons of social engineers (Krombholz et al., 2015).

With the prevalence of both technical, socio-technical, and non-technical research on social engineering, a broad review of the literature can reveal the dominant themes within these aspects of social engineering research. Moreover, performing such an overarching review of cybersecurity-related social engineering research will help us understand the breadth of areas it encompasses. Next, we discuss the research methodology used in our study.

RESEARCH METHODOLOGY

We used a multi-step approach in examining social engineering research to unravel the research themes. In this section, we detail our data collection and data analysis methods for this study.

Data Collection

As a first step in our research, we collected data on published research from ISI's Web of Science (WOS) databases. We included a wide array of WOS databases as cybersecurity-related social engineering research has been done with both technical, socio-techical, and non-technical focus. Moreover, conference proceedings being a popular avenue for some of the disciplines in cybersecurity, we decided to include that WOS database. Hence, the following WOS databases were used: Science Citation Index Expanded, Conference Proceedings Citation Index - Science (CPCI-S), Social Science Citation Index (SSCI), and Emerging Sources Citation Index (ESCI).

We used the search term "Social Engineering," yielding 1,497 unique scholarly articles. Next, we reviewed the results and selected cybersecurity-related scholarly articles resulting in 819 articles. We downloaded the author details, title, abstract, and the keywords provided in the databases for these articles. Next, we used co-word analysis, a traditional bibliometric technique to identify perspectives on the research themes.

Co-word Analysis

Co-word analysis is a well-established technique that analyses co-occurring words in articles and creates maps of relationship networks among the words revealing the emergent themes in the research area. Since the primary aim of this research study is to unravel the research themes in cybersecurity-related social engineering research, we chose to perform a co-occurrence analysis of keywords in published research articles. Author-supplied keywords are known to capture the theme of the research study succinctly.

Before performing the co-word analysis, we cleaned the dataset to ensure keyword usage consistency. For example, some papers have used "cybersecurity" as a keyword, while others have used "cyber security" as the keyword in the manuscript. Likewise, there were other keywords where different articles used slightly different ways to represent the keywords. Some of such variations include "attack" and "attacks," "cyber crime" and "cybercrime", etc.

We used VOSviewer, a popular bibliometric tool, to perform the co-occurrence of keywords analysis and to create a visual map of the relationships between keywords and cluster themes. Van Eck and Waltman (2010) outlined the Visualization of Similarities (VOS) mapping method for constructing and visualizing bibliometric networks such as co-word. VOSviewer has been used by researchers to understand the research domain by analyzing the co-occurrence of article keywords. For example, Mangalaraj et al. (2022) explore SOX-related information systems research themes and analyze the literature through co-word analysis to unravel the research's eight overarching themes. During the analysis with VOSviewer, we specified at least five times a keyword to occur before it is included in further analysis. In total, 70 keywords met this criterion, so we proceed further with the analysis and interpretation of the data.

RESULTS

Table 2 shows the distribution of articles by journals with more count more than 3. As can be seen, there are many publication outlets for research on cybersecurity-related social engineering. Of the 819 articles considered for this review, 462 were published in conference proceedings.

Table 2. Distribution of Articles by outlet with count > 3

PUBLICATION OUTLET	NUMBER OF ARTICLES
IEEE Access	27
Computers & Security	23
Applied Sciences	9
International Journal of Computer Science and Network Security	9
Computers In Human Behavior	8
International Journal of Advanced Computer Science and Applications	7
Electronics	6
Information	6
Security And Communication Networks	6
Security And Privacy	6
Information And Computer Security	5
Expert Systems with Applications	4
Frontiers In Psychology	4
Future Internet	4

Figure 1 reveals the number of articles on cybersecurity-related social engineering appearing in a year. Interestingly, the very first research on social engineering related to computing appeared in the late 1990s. However, cybersecurity-related social engineering research has seen exponential growth in recent years.

Figure 1. Number of articles on Social Engineering per year

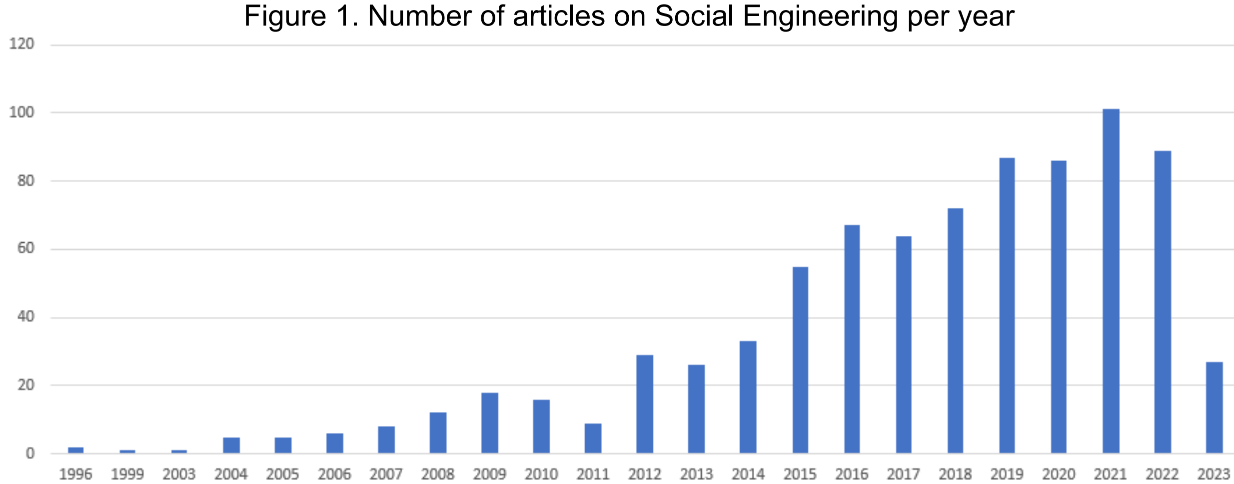


Figure 2 presents the keyword network map of research on cybersecurity-related social engineering. Each keyword is represented in the network map with its relationship with other keywords. Moreover, the keyword map also presents the dominant themes in using the co-occurrence of article keywords by depicting them in different colors. As seen in figure 2, there are 7 identified thematic clusters in the social engineering research.

Figure 2. Keyword network map for the period 1996-2023

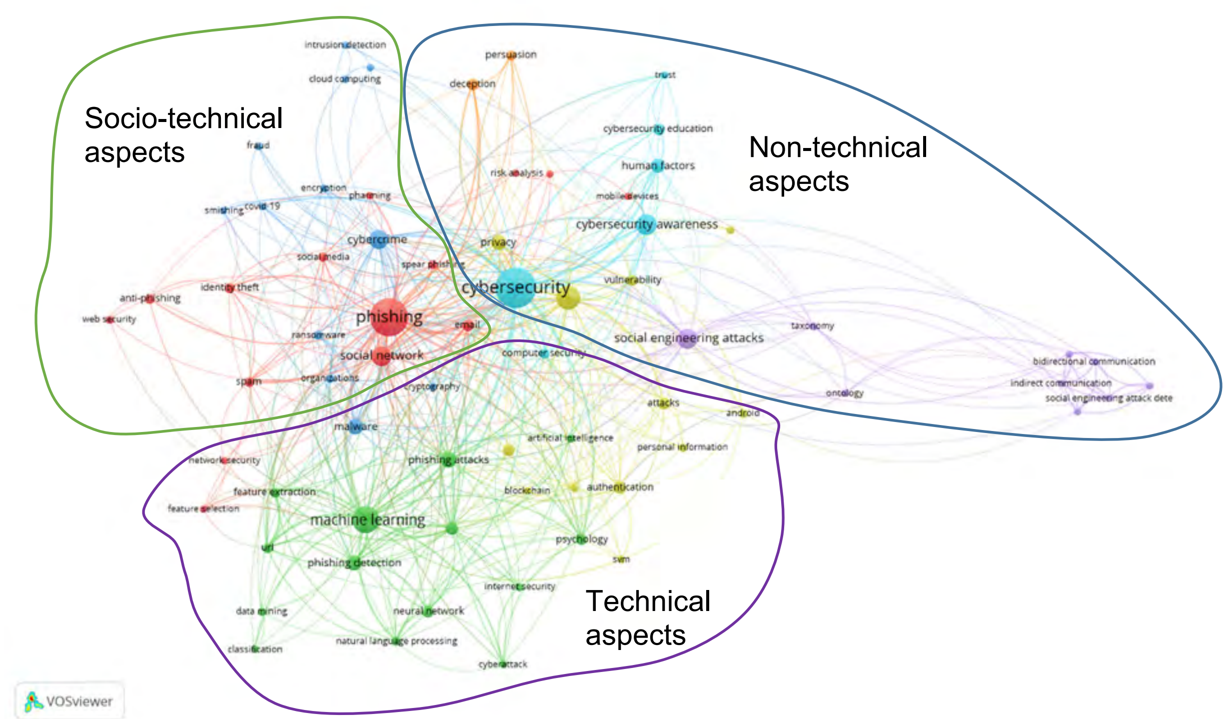


Table 3 shows cluster themes and the associated keywords based on the VOSviewer keyword co-occurrence analysis. We have also indicated the technical, non-technical, and socio-technical aspects. Authors independently examined the keywords in a given theme and collectively named these clusters, as shown in the table below.

Table 3. Co-occurrence of keywords and cluster themes		
CLUSTER	KEYWORDS	CLUSTER-THEME
Socio-technical aspects		

Cluster 1	Anti-phishing, email, feature selection, identity theft, mobile devices, network security, pharming, phishing, risk analysis, social media, social network, spam, spear phishing, vulnerability analysis, web security	Social engineering risks and mitigation
Cluster 2	Cloud computing, COVID-19, cryptography, cybercrime, encryption, fraud, Internet of things, intrusion detection, malware, organizations, ransomware, smishing	Social engineering-related threats and protections
Technical aspects		
Cluster 3	Artificial intelligence, classification, cyberattack, data mining, deep learning, feature extraction, internet security, machine learning, natural language processing, neural networks, phishing attacks, phishing detection, psychology, and URL	Phishing detection mechanisms
Cluster 4	Advanced persistent threat, android, attacks, authentication, blockchain, Internet, penetration testing, personal information, privacy, security, svm, vulnerability	Evolving threats
Non-technical aspects		
Cluster 5	Bidirectional communication, indirect communication, ontology, social engineering attack detection model, social engineering attack examples, social engineering attack framework, social engineering attacks, taxonomy, unidirectional communication	Social engineering attacks overview and communication
Cluster 6	Computer security, cybersecurity, cybersecurity education, cybersecurity education, human factors, trust	Cybersecurity education and awareness
Cluster 7	Deception, persuasion	Deception and persuasion

DISCUSSION AND IMPLICATIONS

In this section, we discuss the major themes present in the research on social engineering. It is interesting to note that of the seven streams of research identified in our study, three of them deal with non-technical areas of social engineering-related research. Moreover, this diversity in the research shows that social engineering is a multi-faceted phenomenon that cannot be solved with technical solutions.

Cluster 1: Social Engineering Risks and Mitigations

The first cluster dealt with the phishing aspects of social engineering and the mechanisms used, such as spear phishing. Since social media is one of the mechanisms through which phishing attempts are made, that, too, found a place in this cluster. This cluster also had risk analysis and vulnerability analysis.

Research on social media in this theme included studies on how social media can provide data for social engineering (McNealy, 2022). It can serve as a medium through which users fall victim to social engineering attacks (Frauenstein & Flowerday, 2016). Identity theft is a fast-growing threat affecting individuals, and researchers have examined the social engineering aspects of perpetrating such threats by compromising personal information (Al-Charchafchi et al., 2020).

Cluster 2: Cyber Threats and Protections

The second cluster pertains to cybersecurity threats such as malware, ransomware, and cloud computing and their protection, such as cryptography, introduction detection, and encryption. This cluster also had fraud and cybercrime. Interestingly, this cluster also had the keyword COVID-19.

The presence of malware and ransomware in this cluster shows the far-reaching effects of social engineering, as it is not limited to just phishing alone. Some ransomware attacks originate as social engineering attacks (Moussaileb et al., 2020) and bypass many technical solutions by leveraging social engineering methods (Moore, 2016). Another area of research on this theme is fraud and cybercrime. For example, the recent pandemic was exploited by fraudsters to design requests around the crisis affecting many to carry out social engineering attacks (Taodang & Gundur, 2022). Moreover, the pandemic shifted various everyday activities to online, providing more opportunities for social engineering attacks (Venkatesha et al., 2021).

Cluster 3: Phishing Detection

The third cluster had techniques used for detecting social engineering attacks such as phishing. As seen from the keywords in the cluster, many Artificial Intelligence-related techniques such as Machine Learning (ML), Natural Language Processing (NLP), and Deep Learning are used for phishing detection-related research. Researchers have used these AI techniques on various aspects of social engineering.

Artificial intelligence techniques such as supervised and unsupervised learning have been used to detect phishing attempts by analyzing websites' URLs (Uniform resource locator) (Indrasiri et al., 2021). Moreover, these techniques were also used beyond the links in the email to the content of the email itself for spam detection (Shaaban et al., 2022).

Cluster 4: Emerging threats

The fourth cluster dealt with the emerging threats regarding social engineering. For example, this cluster theme had research on threats in the mobile/android, advanced persistent threats, and blockchain areas. Moreover, this cluster also researched privacy and personal information, as social engineering attacks often compromised them.

The advanced persistent threat is an emerging area that is difficult to detect as the attackers combine phishing emails, malware, botnet, and other social engineering techniques to create a series of attacks in one APT attack (Lu et al., 2017). With blockchain becoming popular in recent years, researchers have examined the suitability of it to prevent social engineering attacks through properly recording and controlling them (Fakieh & Akremi, 2022). Mobile devices and corresponding apps are increasingly used as the primary mode for many to access the Internet and are a rich target for social engineering attacks. Researchers have proposed permission models to deal with the need to enhance the privacy of the data these applications have access (Alshehri et al., 2019).

Cluster 5: Social Engineering Attacks Overview and Communication

The fifth cluster deals with the communication aspect of social engineering. This cluster is one of the two cluster themes that were specifically focused on the interaction between the social engineering attacker and the victim. Research in this stream shows the pervasiveness of the non-technical side of social engineering research.
Research in this stream considers the communication aspect between the social engineer, and the target through a medium (Mouton et al., 2016). Researchers proposed social engineering detection models that include the directionality of the communication between the attacker and victim with classifications including bidirectional, unidirectional, and indirect communication (Mouton et al., 2015).

Cluster 6: Cybersecurity Education and Awareness

The sixth cluster identified by the co-word analysis belonged to the human aspects of social engineering and the ways to mitigate that through security education and awareness. Studies underscore the importance of non-technical defenses against social engineering attacks, such as security education, training, and awareness. Hence it is not a surprise that there is a dedicated stream of research on cybersecurity education and awareness.

Researchers have proposed a game to improve social engineering awareness among technology consumers (Olanrewaju & Zakaria, 2015). Aldawood and Skinner (2019), comprehensively reviewed cybersecurity social engineering training and awareness programs. Since human factors such as trust play a major role in social engineering attacks, interventions aimed at warning or awareness about social engineering have also been focus of studies (Junger et al., 2017). Rege et al. (2020) describe developing and implementing awareness and training programs.

Cluster 7: Deception and Persuasion

Interestingly, the last cluster had two keywords, "deception" and "persuasion." Principles in persuasion and deception go into the core of social engineering. While the previous cluster 5 dealt with the communication modalities and their directionality, this cluster focussed on the methods for convincing and deceiving the victim of social engineering attack.
Some examples of studies in this area include persuasion principles such as liking and its effect on clicking on phishing emails (Lawson et al., 2019). Likewise, persuasive principles such as authority are more often used when compared to others (Bullee et al., 2018). For deceiving the victim, the message characteristics influence the susceptibility to social engineering attacks (Algarni, 2019). Hence, this stream of research focuses explicitly on how the attacker tricks the victim in these social engineering attacks.

IMPLICATIONS

Our study found seven dominant themes of research on cybersecurity-related social engineering research, with two themes dealing with the technical aspects and socio-technical aspects each and three dealing with non-technical aspects. It is interesting to note that social engineering itself is considered to be a non-technical means for hackers to gain access (Cullen & Mann, 2008), and the prevalence of technical research in this area shows the important ways organizations are mitigating social engineering threats.

Our study also shows that researchers have focussed more on phishing detection-related social engineering research as it forms a specific theme. Our findings show the importance of phishing in the social engineering research stream. However, now we have emerging forms of social engineering that form part of other cyber-attacks, such as ransomware and advanced persistent threats. Research should focus on detection techniques for such evolving threat landscapes for social engineering research.

It is interesting to note that the keywords deception and persuasion itself form a theme in our analysis. With social engineering relying on tricking the victim, it is no wonder this theme is prominent in the extant research. Moreover, it also indicates that the research on this theme is anchored in many aspects of social engineering attacks, not necessarily with particular social engineering attacks such as phishing alone.

It is also interesting to note that the cluster on communication shows an interesting pattern with many of the communication related keywords (unidirectional, bidirectional, and indirect communication) closely coupled to each other, indicating a cohesive stream of research in this area. Moreover, this theme also has social engineering attack-related keywords. Hence, more research on focussed areas of social engineering, such as phishing, can bridge the chasm otherwise revealed in the keyword network map.

Finally, the cybersecurity education and awareness theme show the importance of these aspects in mitigating the risks of social engineering attacks. As our research reveals, there are emerging threats with social engineering, and the curriculum for these awareness and training programs should be frequently updated to keep up with the development in this area. Just like the frequent updates with anti-virus programs to keep up with the latest attack vectors, training, and awareness programs should be constantly refined to keep up with the zero-day threats in social engineering.

Limitations of the study

Research using the co-occurrence of keywords has certain limitations. For example, the collection of articles that is the source of the analysis influences the study outcomes. Though we have used diverse ISI Web of Science databases, there is the probability that certain publications may not have been indexed in these databases. Including diverse scholarly publications provides credence as the sample used is representative, and the exclusion of other research venues may not be a significant issue.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

We have witnessed an exponential increase in cybersecurity-related social engineering incidents since the start of the century. Though phishing is a common social engineering attack, we have recently seen other emerging areas, such as ransomware, an advanced persistent threat. Hence, research on social engineering too has seen a concurrent increase in focus by the researchers. We used co-word analysis to unravel the dominant cybersecurity-related social engineering research themes. We discussed the seven clusters that emerged from our study and also discussed these study implications. Our study findings can help the academic and practitioner communities better understand social engineering in cybersecurity.

REFERENCES

Abraham, S., & Chengalur-Smith, I. (2010). An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, *32*. https://doi.org/10.1016/j.techsoc.2010.07.001

Abu Zuraiq, A., & Alkasassbeh, M. (2019). Review: Phishing Detection Approaches. 2019 2nd International Conference on New Trends in Computing Sciences(ICTCS),

Al-Charchafchi, A., Manickam, S., & Alqattan, Z. (2020). Threats Against Information Privacy and Security in Social Networks: A Review. Advances In Cyber Security (ACES 2019),

Aldawood, H., & Skinner, G. (2019). Reviewing Cyber Security Social Engineering Training and Awareness Programs-Pitfalls and Ongoing Issues. *Future Internet*, *11*. https://doi.org/10.3390/fi11030073

Algarni, A. (2019). What Message Characteristics Make Social Engineering Successful on Facebook: The Role of Central Route, Peripheral Route, and Perceived Risk. *Information*, *10*. https://doi.org/10.3390/info10060211

Alshehri, A., Marcinek, P., Alzahrani, A., Alshahrani, H., & Fu, H. (2019). PUREDroid: Permission Usage and Risk Estimation for Android Applications. Proceedings of 3rd International Conference on Information System and Data Mining (ICISDM 2019),

Asiri, S., Xiao, Y., Alzahrani, S., Li, S., & Li, T. (2023). A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks. *IEEE Access*, *11*. https://doi.org/10.1109/ACCESS.2023.3237798

Bullee, J., Montoya, L., Pieters, W., Junger, M., & Hartel, P. (2018). On the anatomy of social engineering attacksA literature-based dissection of successful attacks. *Journal Of Investigative Psychology And Offender Profiling*, *15*. https://doi.org/10.1002/jip.1482

Cullen, A., & Mann, I. (2008). Hacking the Human: Countering the Socially Engineered Attack. Proceedings of the 7th European Conference On Information Warfare And Security,

Edwards, M., Larson, R., Green, B., Rashid, A., & Baron, A. (2017). Panning for gold: Automatically analysing online social engineering attack surfaces. *Computers & Security*, *69*. https://doi.org/10.1016/j.cose.2016.12.013

Fakieh, A., & Akremi, A. (2022). An Effective Blockchain-Based Defense Model for Organizations against Vishing Attacks. *Applied Sciences*.

Frauenstein, E., & Flowerday, S. (2016). Social Network Phishing: Becoming Habituated to Clicks and Ignorant to Threats? 2016 Information Security For South Africa - Proceedings of the 2016 ISSA Conference,

Gupta, S., Singhal, A., & Kapoor, A. (2016). A Literature Survey on Social Engineering Attacks: Phishing Attack. 2016 IEEE International Conference on Computing, Communication and Automation (ICCCA),

Hatfield, J. (2018). Social engineering in cybersecurity: The evolution of a concept. *Computers & Security*, *73*. https://doi.org/10.1016/j.cose.2017.10.008

Hijji, M., & Alam, G. (2021). A Multivocal Literature Review on Growing Social Engineering Based Cyber-Attacks/Threats During the COVID-19 Pandemic: Challenges and Prospective Solutions. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2020.3048839

Indrasiri, P., Halgamuge, M., & Mohammad, A. (2021). Robust Ensemble Machine Learning Model for Filtering Phishing URLs: Expandable Random Gradient Stacked Voting Classifier (ERG-SVC). *IEEE Access*.

Junger, M., Montoya, L., & Overink, F. (2017). Priming and warnings are not effective to prevent social engineering attacks. *Computers in Human Behavior*.

Krombholz, K., Hobel, H., Huber, M., & Weippl, E. (2015). Advanced social engineering attacks. *Journal of Information Security And ApplicatiONS*, *22*. https://doi.org/10.1016/j.jisa.2014.09.005

Lawson, P., Crowson, A., & Mayhorn, C. (2019). Baiting the Hook: Exploring the Interaction of Personality and Persuasion Tactics in Email Phishing Attacks. Proceedings of The 20th Congress of The International Ergonomics Association (IEA 2018), Vol V: Human Simulation And Virtual Environments, Work With Computing Systems (WWCS), Process Control,

Lu, J., Zhang, X., Wang, J., & Ying, L. (2017). APT Traffic Detection Based on time transform. 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS),

Luo, X., Brody, R., Seazzu, A., & Burd, S. (2011). Social Engineering: The Neglected Human Factor for Information Security Management. *Information Resources Management Journal*, *24*. https://doi.org/10.4018/irmj.2011070101

Malatji, M., Marnewick, A., & von Solms, S. (2020). Validation of a socio-technical management process for optimising cybersecurity practices. *Computers & Security*, *95*, 101846.

Mangalaraj, G., Singh, A., & Taneja, A. (2022). Probing the Past to Guide the Future IT Regulation Research: Topic Modeling and Co-word Analysis of SOX-IS Research. *Information Systems Management*, 1-14.

Mashtalyar, N., Ntaganzwa, U., Santos, T., Hakak, S., & Ray, S. (2021). Social Engineering Attacks: Recent Advances and Challenges. HCI for Cybersecurity, Privacy and Trust (HCI-CPT 2021),

Matyokurehwa, K., Rudhumbu, N., Gombiro, C., & Chipfumbu-Kangara, C. (2022). Enhanced social engineering framework mitigating against social engineering attacks in higher education. *Security And Privacy*, *5*. https://doi.org/10.1002/spy2.237

McNealy, J. (2022). Platforms as phish farms: Deceptive social engineering at scale. *New Media & Society*, *24*. https://doi.org/10.1177/14614448221099228

Moore, C. (2016). Detecting Ransomware with Honeypot techniques. 2016 Cybersecurity and Cyberforensics Conference (CCC),

Moussaileb, R., Cuppens, N., Lanet, J., & Le Bouder, H. (2020). Ransomware Network Traffic Analysis for Pre-encryption Alert. Foundations And Practice Of Security, FPS 2019,

Mouton, F., Leenen, L., & Venter, H. (2015). Social Engineering Attack Detection Model: SEADMv2. 2015 International Conference on Cyberworlds (CW),

Mouton, F., Leenen, L., & Venter, H. (2016). Social engineering attack examples, templates and scenarios. *Computers & Security*.

Olanrewaju, A., & Zakaria, N. (2015). Social Engineering Awareness Game (SEAG): An Empirical Evaluation of Using Game Towards Improving Information Security Awareness. Proceedings of the 5th International Conference on Computing & Informatics

Rege, A., Nguyen, T., & Bleiman, R. (2020). A social engineering awareness and training workshop for STEM students and practitioners. 2020 9TH IEEE Integrated Stem Education Conference (ISEC 2020),

Salahdine, F., & Kaabouch, N. (2019). Social Engineering Attacks: A Survey. *Future Internet*, *11*. https://doi.org/10.3390/fi11040089

Shaaban, M., Hassan, Y., & Guirguis, S. (2022). Deep convolutional forest: a dynamic deep ensemble approach for spam detection in text. *Complex & Intelligent Systems*, *8*. https://doi.org/10.1007/s40747-022-00741-6

Siadati, H., Nguyen, T., Gupta, P., Jakobsson, M., & Memon, N. (2017). Mind your SMSes: Mitigating social engineering in second factor authentication. *Computers & Security*, *65*. https://doi.org/10.1016/j.cose.2016.09.009

Siddiqi, M., Pak, W., & Siddiqi, M. (2022). A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures. *Applied Sciences*, *12*. https://doi.org/10.3390/app12126042

Steinmetz, K., Pimentel, A., & Goe, W. (2021). Performing social engineering: A qualitative study of information security deceptions. *Computers in Human Behavior*, *124*. https://doi.org/10.1016/j.chb.2021.106930

Syafitri, W., Shukur, Z., Mokhtar, U., Sulaiman, R., & Ibrahim, M. (2022). Social Engineering Attacks Prevention: A Systematic Literature Review. *IEEE Access*, *10*. https://doi.org/10.1109/ACCESS.2022.3162594

Taodang, D., & Gundur, R. (2022). How Frauds in Times of Crisis Target People. *Victims & Offenders*. https://doi.org/10.1080/15564886.2022.2043968

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *scientometrics*, *84*(2), 523-538.

Venkatesha, S., Reddy, K. R., & Chandavarkar, B. (2021). Social engineering attacks during the COVID-19 pandemic. *SN computer science*, *2*, 1-9.

Yasin, A., Fatima, R., Liu, L., Yasin, A., & Wang, J. (2019). Contemplating social engineering studies and attack scenarios: A review study. *Security and Privacy*, *2*. https://doi.org/10.1002/spy2.73

DECISION SCIENCES INSTITUTE

The Power of Success: How High-Performing Organizations Navigate Major Transformations Without Sacrificing Credibility

Majid Majzoubi
York University
majzoubi@yorku.ca

ABSTRACT

Our research investigates the relationship between high-performing organizations, their ability to make major transformations, and the maintenance of their credibility. The cognitive economy theory provides a framework for suggesting that a company's performance plays a significant role in its ability to implement changes without negatively impacting its credibility. Superior performing organizations, backed by their increased credibility, are found to have more room to navigate substantial transformations. We anticipate that high-dynamism industries are more lenient towards such transformations. Our hypotheses are put to the test through the lens of coverage decisions made by security analysts concerning U.S. publicly traded companies from 1997 to 2018. Natural Language Processing methods have been harnessed to develop a novel method for tracking substantial transformations within organizations.

KEYWORDS: Organizational transformation, Firm credibility, Company preformance

DECISION SCIENCES INSTITUTE

The role of leadership styles, organizational culture, and knowledge management in higher education institutions (HEI)

Armel Djangone
Dakota State University
Email: armel.djangone@trojans.dsu.edu

Omar El-Gayar
Dakota State University
Email: omar.el-gayar@dsu.edu

ABSTRACT

Knowledge management (KM) strategies adopted by universities are either inadequate or inconsistent. Minimal research has explored the link between KM enablers and the effective implementation of KM in the context of HEI. This empirical study aims to investigate the relationship between transformational leadership, transactional leadership, organizational culture, KM effectiveness and organizational performance. The research employs structural equation modeling using Partial Least Square (SEM-PLS) to study the underlying relationships. Key findings suggest that organizational culture and transformational leadership contributed to KM effectiveness and that KM is essential for improving the performance of higher education institutions.

KEYWORDS: Knowledge management, Higher education, Transactional leadership, Transformational leadership, Organizational culture, Partial Least Square

INTRODUCTION

Higher education institutions (HEI) in the United States have previously relied on conventional economic models primarily enrolling a balanced number of tuition-paying students and receiving public and private funding from states, federal and other stakeholders (Wiley University Services, 2021). However, this sector has suffered significant changes over the past decades such as the major transformations in the technological environment (Fernández-López et al., 2018), the complex social demands in HEIs (Quarchioni et al., 2022), and the rapid development of the global marketplace in which increasingly HEIs operate (Quarchioni et al., 2022). These are just a few of the challenges these organizations are facing over time. Additionally, the decrease in public funding for colleges and universities and a growing number of normative calls for transparency, competitiveness, and quality (Quarchioni et al., 2022) have put increasing pressure on HEIs. This burden on HEIs throughout the nation requires them to do more with less public funding while raising the number of graduate students and managing the cost of operation.

The shift from natural resources to the knowledge era in the corporate setting has positioned knowledge as an essential asset to gain a sustainable competitive advantage (Si Xue, 2017), consequently establishing knowledge management practice as a critical resource for organizations. Scholars (Nawaz et al., 2014; Rodríguez-Gómez & Gairín, 2015) increasingly discuss KM use in HEIs. The challenge for HEIs to remain competitive allows researchers to show the great potential of KM for HEIs to use it to improve research activities student learning outcomes, and enrich core curriculum processes (Khakpour, 2015), and improve the overall performance (Ahmad et al., 2015; Iqbal et al., 2019). It is thus paramount to understand the KM enablers and processes that maximize the capabilities of HEI to manage their knowledge assets.

HEI thus need to know how to properly manage their knowledge assets by effectively expanding, disseminating, and exploiting them and recognizing KM success factors. Several studies investigated the critical success factors (CSF) of KM implementation in business organizations (Aldulaimi, 2015; Dhamdhere, 2015; Razmerita et al., 2016). They refer to factors, such as leadership and culture that facilitate KM processes or activities (Areed et al., 2021). Leadership is an essential and a fundamental aspect of a business context (Millar et al., 2018). In effect, successes and failures depend on the leader's integrity, team diversity management, and the organization's knowledge as well as on a leader's ability to grow and exploit their team's potential and, thus, overall organizational capability. According to Baker et al. (2002), organizational culture is the informal, nonmaterial, interpersonal, and moral bases of cooperation and commitment that are more important than the formal, material, and instrumental controls. Organizational culture works as the normative glue that allows for coordination and stability.

Despite the significant importance of KM in organizations in general, exploring the interaction of KM, leadership styles, and organizational performance particularly in the context of HEI remains understudies (Ahmady et al., 2016; Veer Ramjeawon & Rowley, 2017). Such context exhibits distinguishing characteristics such as shared governance that may reflect on such interaction. Accordingly, this research explores the relationship between leadership styles, organizational culture, KM, and organizational performance in the context of HEI. This research emphasizes leadership styles corresponding to 'middle management' in other domains, e.g., department chairs, deans, and program directors and their organizational context, e.g., school, department, or unit. Specifically, the research aims to answer the following research questions:

1. What is the relationship between transformational leadership, transactional leadership and knowledge management in the context of HEI?
2. How does KM affect the performance in the context of HEI?
3. What is the role of organizational culture in relation to knowledge management effectiveness and leadership style in the context of HEI?

This study provides a framework for understanding the type of leadership and its relation to organizational culture and knowledge management effectiveness in the context of HEI. In essence, this study answers the call of Iqbal et al. (2019) by focusing on the relationship between knowledge management and organizational performance, in addition to the relationship between organizational leadership and organizational culture on knowledge management. The next section of the paper summarizes the literature and highlights the research gap. The research model, hypotheses, and research methodology follow this. The remaining sections present the results, discuss key findings, and conclude the paper.

LITERATURE REVIEW

Addressing challenges confronting HEI, Sunalai and Beyerlein (2015) suggest that HEI use knowledge to acquire and sustain competitive advantage. Across the past decade, there has been a growing interest in the impact of KM on organizational performance (Sunalai & Beyerlein, 2015). Higher education institutions are inherently knowledge-intensive organizations involved in learning, creation, and contribution to the public good. However, Si Xue (2017) warns that the benefit of knowledge is only perceived when correctly managed. Consequently, Rowley (2000) investigated the readiness of HEI for KM. They concluded that the core activities of HEI are associated with knowledge creation, dissemination, and learning and that HEI are inherently knowledge intensive. Further, Kidwell et al. (2000) examined the applicability of corporate KM to HEI. They outlined basic concepts of corporate KM and considered trends to determine whether HEI were ready to embrace corporate KM. The study suggests that HEI have substantial chances to apply KM practices to support the core activities of their mission. External forces such as market competition, internalization, and the global business environment led to new ways of understanding the role and functions of HEI and the increased potential of KM in such a context (Dhamdhere, 2015; Rowley, 2000). According to Kidwell et al. (2000) KM can help address such challenges resulting in improved decision-making skills, enhanced academic and administrative services, and decreased expenditures.

KM can positively impact the performance of HEI by promoting the services and processes (Ahmad et al., 2020; Ahmad et al., 2015; Iqbal et al., 2019). For instance, using 300 responses from teaching staff, Ahmad et al. (2020) studied the relationship between the KM processes and job performance in HEI in Pakistan. The study found that seven KM processes, including knowledge creation, sharing, and utilization, positively influence job performance in the context of HEI, and job performance significantly influences organizational performance. Their effort is essential to the current study as it demonstrates that KM processes are also critical to the performance of educational institutions. Ahmad et al. (2015) studied the effects of KM components such as information technology, organizational learning, and knowledge on the performance of HEI. The study used a sample size of 113 faculty and administrative staff from 26 universities. The findings indicate that KM significantly affects the organizational performance of universities and colleges. Similarly, Sahibzada et al. (2020) investigated the multifaceted correlations between KM processes (acquisition, sharing and utilization), knowledge worker satisfaction, and organizational of HEIs. Using a sample of 238 academics and administrative staffs the results support that KM processes significantly affect knowledge worker satisfaction which in term improve the performance of HEI.

While KM is broadly recognized as an essential strategic resource for the performance of HEI, various factors influence its effectiveness, including leadership and culture. Ather and Awan (2021) examined the effect of KM practices and transformational and transactional leadership on university teachers' performance. Data collected from 260 teaching faculty members were analyzed through regression analysis. The study demonstrated the positive effects of KM practices on university teacher performance. Moreover, the study suggested that transformational and transactional leadership are essential to KM practices, and that transformational leadership performs better than transactional leadership. This research is relevant to our study as it investigates the effects of transformational and transactional leadership on KM practices. Iqbal et al. (2019) empirically investigated the effects of KM enablers on KM processes in research universities. Using a sample of 217 academic and administrative staff from research universities in Pakistan, the findings suggest that KM enablers such as culture, leadership, and incentives facilitate KM processes. Furthermore, the research found that effective implementation of KM processes (knowledge acquisition, sharing, and utilization) is essential to the performance of HEI. Their research is crucial to our current study as it provides insights for understanding the interplay between leadership, culture, knowledge management effectiveness, and organization performance in the context of HEI. Ugwu and Okore (2020) reported a direct and positive effect of transformational and transactional leadership on KM effectiveness. They further reported that transformational leadership impacted KM more than transactional leadership. This study appears relevant to our research as it highlights the type of leadership style that impacts KME and demonstrates that transformational leadership was more influential than transactional leadership. Akhavan et al. (2014) studied the role of leadership style and organizational culture in the effectiveness of KM. A sample size of 224 employees from four research centers was analyzed where the result showed a positive impact of culture and leadership on KM implementation within research centers. This study is relevant to our research as it provides ways to understand the important factors that facilitate KM processes in research centers. Alshahrani (2018) conducted a similar study to understand the effects of various factors on KM processes in HEI using a semi-structure interview. The analysis of the 25 participants responses showed that various factors including organization culture, and leadership influence the implementation of KM practices.

Despite the richness of research exploring the influence of various factors on KM implementation, the interplay between transformational and transactional leadership, organization culture, knowledge management, and organizational performance remains understudied in the context of HEI. While the literature shows much interest in organizational knowledge management activities, little attention was given to explore the collective role of KM in the context of HEI and roles of various critical factors that facilitate its proper implementation in such context. Recognizing the complexity of the educational business environment and its distinguishing characteristics, this study aims to complement the existing literature by investigating the joint role of organizational culture, leadership styles, KM, and organizational performance in the context of HEI.

RESEARCH MODEL

This study examines the relationship between leadership styles, organizational culture, and KM processes. Additionally, the study investigates the direct influence of KM processes on organizational performance (OP), and the mediating effect of organizational culture in the relationship between KM and organizational leadership in HEI. Figure 1 presents the research model adopted for this study.

Effective leadership plays a significant role in ensuring an organizational initiative’s success. Two styles to leadership (transformational and transactional leadership) were adopted from Deichmann and Stam (2015) to demonstrate the role of leadership in KM effectiveness. Transformational leadership (TFL) emphasizes motivation and inspiration. They motivate subordinates to achieve performance beyond expectations, including components such as idealized influence, inspirational motivation, intellectual stimulation, and individualized consideration (Avolio & Bass, 1995). Researchers have observed a positive relationship between transformational leadership and knowledge-creation processes (Avolio et al., 1999). It is demonstrated that transformational leaders play a critical role in developing interactions and relationships needed for creating social capital and opportunities for employees to explore new ideas and knowledge (Sayyadi, 2019). For instance, idealized influence, inspirational motivation, and intellectual stimulation, which inspire followers to take risks and generate innovative solutions, are essential components of transformational leadership that support KM (Sayyadi, 2019). Given these arguments, the following research hypothesis can be deduced:

> H1. Transformational leadership has a significantly positive effect on the KM effectiveness in the context of HEI.

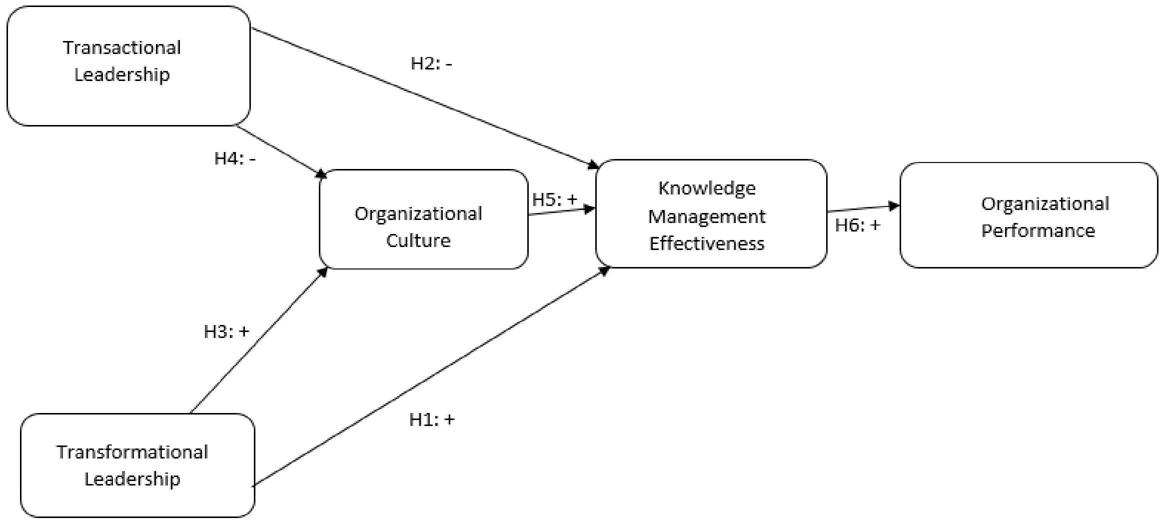


Figure 1: research model - Plus and minus symbols represent positive and negative relationships, respectively.

Transactional leadership (TSL) requires a relationship between leaders and their followers. In this relation, followers receive rewards and resources for fulfilling a leader's needs. These leaders are accountable for maintaining routine tasks by managing individual performance and relying on self-motivated people. TSL has three main components: contingent reward, active management-by-exception, and passive management-by-exception. TSL can improve the efficiency of learning in an organization by emphasizing existing values and routines and focusing on increasing efficiency in current practices, which enables transactional leaders to foster rule-based ways of doing things (Bass & Avolio, 1994). Transactional leaders also provide organizational members with formal systems and training programs that disseminate existing learning to guide future actions and decisions, which participate in KM creation, sharing, and application. Based on the above, the following research hypothesis can be deduced:

> H2. Transactional leadership has a significantly positive effect on the KM effectiveness in the context of HEI.

Organizational culture has emerged as an issue of vital interest within the business community for the last ten years (Tierney & Lechuga, 2004). Organizational culture creates an environment in which every employee attempt to achieve the goal set by the organization. Therefore, understanding culture is essential for the employees, the stakeholders, and the organization. Leadership is one of the most influential factors in an organization's culture (Nam Nguyen & Mohamed, 2011). During the organization formation process, a company's founder creates an organization that reflects the values and beliefs they believe are necessary and good for the organization. In this sense, the founder creates and shapes the cultural traits of their organizations (Nam Nguyen & Mohamed, 2011). Therefore, the leadership approach can serve as a critical framework for creating and managing organizational culture. Organizational culture is defined as a pattern of fundamental assumptions invented, discovered, or developed by a given group as it learns to cope with the problems of external adaptation and internal integration (Acar, 2012). Therefore, the following two hypotheses can be deducted for both transactional and transformational leadership:

> H3. Transformational leadership style has a significantly positive effect on the organizational culture.

H4. Transactional leadership style has a significantly positive effect on the organizational culture.

The relationships between transformational or transactional leadership styles and KM depend upon various organizational culture types (Shirini & Xenikou, 2022). Specifically, Shirini and Xenikou (2022) found that organizational culture mediates the relationship between transformational leadership and organizational learning and KM.

Seyedyousefi et al. (2016) state that organizational culture can influence KM effectiveness. They further stated that for successful KM, organizations require an environment with a continuous learning culture, and it must occur at all levels of the organization. In a learning culture, employees seek problems, and they will be encouraged to learn. Through learning tools such as implementation learning, they will reinforce KM. By strengthening the learning culture, the capacity of knowledge management implementation will be increased (Seyedyousefi et al., 2016). Thus, we can deduct the following research hypothesis:

H5. Organizational culture has a significantly positive effect on the KM effectiveness of in the context of HEI.

Our research model captures three processes to KM effectiveness often referred to in the literature (Chen & Nonaka, 2022; Chiu & Chen, 2016). These include knowledge acquisition, sharing, and utilization. Some researchers argued that knowledge acquisition, sharing, and utilization improve organizational collective learning and decision-making and enhance productivity and profitability. Consequently, studies revealed a significant, optimistic, and direct association between KM processes and OP (Ahmad et al., 2020; Chiu & Chen, 2016). Ahmady et al. (2016) also confirmed a positive and direct association between KM processes and university performance. Organizational performance has been one of the most investigated variables to measure organizational success (Iqbal et al., 2019). In this research, organizational performance refers to perceptions of performance metrics commonly associated with HEIs such as student's satisfaction, curriculum development, school placement rates, and student success rate (Ahmad et al., 2015; Iqbal et al., 2019). Several recent studies suggested that KM practices can improve the performance of educational institutions. Al-Qarioti (2015) found that KM appears to significantly impact the performance of HEI, providing further evidence of the impacts of KM on the performance of higher education institutions. Consequently, the following hypothesis can be extrapolated:

H6. KM effectiveness has a positive direct impact on organization performance in the context of HEI.

METHODOLOGY

Measures

This study employs two styles of organizational leadership (transactional and transformational), four dimensions of organizational culture (mission, adaptability, involvement, and consistency), three dimensions of KM effectiveness (acquisition, sharing, and utilization), and one dimension of organizational performance (Appendix A). The measures reflect the organizational members' perceptions of the construct under consideration. Items measuring the constructs of KM effectiveness and educational institutions were derived from the study of (Ahmad et al., 2015). This study has adopted all items for knowledge processes. A total of 16 items were adopted. Items measuring leadership style constructs were derived from the Multifactor Leadership Questionnaire (MLQ-5X) developed by Avolio and Bass (1995). This study received permission to use all 8-item questions for the transactional leadership style and 20-item questions for the transformational leadership style. Items capturing organizational culture's properties - mission, adaptability, involvement, and consistency were derived from the Denison Organizational Culture Survey (DOCS). A total of 12 items were modified from Fey and Denison (2003).

Data collection

An online survey was administered over a six-month period via electronic mail and social media through Facebook and LinkedIn. The respondents included faculty and administrative leaders (department chairs, deans, program directors) employed at colleges and universities across the United States on five constructs: transformational leadership, transactional leadership, organizational culture, knowledge management effectiveness, and organizational performance. The survey instrument (Appendix A) was comprised of 74-item questions and included three parts: 1) demographic, such as gender, age, and current position, 2) constructs, the first item questions for organizational performance are presented, followed by the item questions for KM effectiveness, item questions for leadership style, and item questions for organizational culture, and 3) open-ended questions for questions where explanations and details were required. Respondents are either an individual in a non-administrative role, e.g., faculty (using the 'rater' form) or a leader (using the 'leader' form) is in a position to comment on the flow of knowledge around their respective organization, e.g., school or department.

Data Analysis

Partial Least Squares Structural Equation Modeling (PLS-SEM) was used to determine causal relationships due to its ability towards exploratory research in the social sciences and its fortitude in estimating causative relationships between constructs (Hair et al., 2019). In addition, PLS-SEM is suitable for handling smaller sample sizes and makes no assumptions regarding data distribution (Hair et al., 2019). SmartPLS version 3.3.3 was used for PLS-SEM model estimation for its acceptance among the academic community and for its efficiency in assessment (Wong, 2019).

Since the survey instrument contains many indicators, items within constructs were parceled prior to hypotheses testing (Lee, 2018). A total of 14 parcels were used as indicators for the hypothesis testing as suggested by Lee (2018) and summarized in Table 1. Using the domain representative approach (Kishton & Widman, 1994), each parcel represents the larger domain: four for transformational leadership, three for transactional leadership, and four for KM effectiveness.

Table 1 Summary of parcels, items, and constructs

Constructs	Parcels	Items
Transformational leadership	Transfo1	IA1, IB1, IM1, IS1, IC1
	Transfo2	IA2, IB2, IM2, IS2, IC2
	Transfo3	IA3, IB3, IM3, IS3, IC3
	Transfo4	IA4, IB4, IM4, IS4, IC4
Transactional leadership	Transac1	CR1, MA1, CR1
	Transac2	CR2, MA2, CR2
	Transac3	CR3, MA3
Organizational culture	OC1	CI1, CA2, CM1, CC1
	OC2	CI2, CA2, CM2, CC2
	OC3	CI3, CA3, CM3, CC3
Knowledge management effectiveness	KME1	KA1, KS1, KU1, KA5
	KME2	KA2, KS2, KU2, KS5
	KME3	KA3, KS3, KU3, KU5
	KME4	KA4, KS4, KU4, KA6
	Total	14

RESULTS

A total of 250 responses were collected resulting in 136 valid observations were collected from the faculty and their leaders employed at US universities and colleges. The general demographics of this research, such as age, gender, and position, are described in Table 2. The target sample group of this research was composed of three types of institution, which included public, private not-for-profit, and private for-profit. Regarding the type of institution, 85% were collected from public schools. Private for-profit and not-for-profit represented about 15%. Regarding gender, 63% of respondents were female and 35%were male. Faculty members represented 79% of the sample.

Table 2 Demographic information

Variables	Values	N	%
	Gender		
Gender	Male	47	35%
	Female	86	63%
	Other	1	1%
	I prefer not to say	2	1%
	Age		
Age	Under 31 years	6	46%
	31 to 35 years	5	4%
	36 to 40 years	20	15%
	41 to 45 years	21	15%
	46 to 50 years	24	18%
	Over 60 years	58	43%
	Respondent's Position		
Respondents' position at the institutions	Faculty member	108	79%
	Assistant Director	1	1%
	Director	8	6%
	Associate Dean	3	2%
	Dean	3	2%
	Senior Manager	6	4%
	Department chair	7	5%
	Working knowledge of Information Systems or Data Management		
Working knowledge of information Systems or Data Management	Familiar	14	10%
	Advanced Beginner	21	16%
	Competent	37	27%
	Expert	13	10%
	Novice	29	21%
	Proficient	21	16%
	Worked in current area/position		
Time spent at current position	Less than one year	10	7%
	More than one and less than 2	7	5%
	More than 2 and less than 3	6	4%
	More than 3 and less than 4	10	7%
	More than 4 and less than 5	3	2%
	More than 5 years	99	73%
	Worked for institute		
Time spent at the institution	Less than 1 year	7	5%
	More than 1 and less than 2	9	7%
	More than 2 and less than 3	10	7%
	More than 3 and less than 4	8	6%
	More than 4 and less than 5	5	4%
	More than 5 years	96	71%
	Type of institution		
	Public	115	85%

Type of institution	Private, not-for-profit	15	11%
	Private, for-profit	6	4%
	Highest level of degrees awarded by your institution		
	Associate	54	40%
	Bachelor	10	7%
	Masters	17	13%
Highest degree conferred by the institutions	Doctoral	51	38%
	Professional (e.g., MD, JD, DDS)	4	2%

Figure 2 represents the research model post-analysis using SmartPLS. Displayed are all indicators (yellow) and loadings for each latent variable construct (white). Arrows pointing to the indicators illustrate a reflective model (Wong, 2019). The indicators and associated constructs represent the measurement model, whereas the exogenous and endogenous variables and their relationships represent the structural model. Endogenous variables are identified with their corresponding composite reliability. The following sections present the test results of bot0.h models.

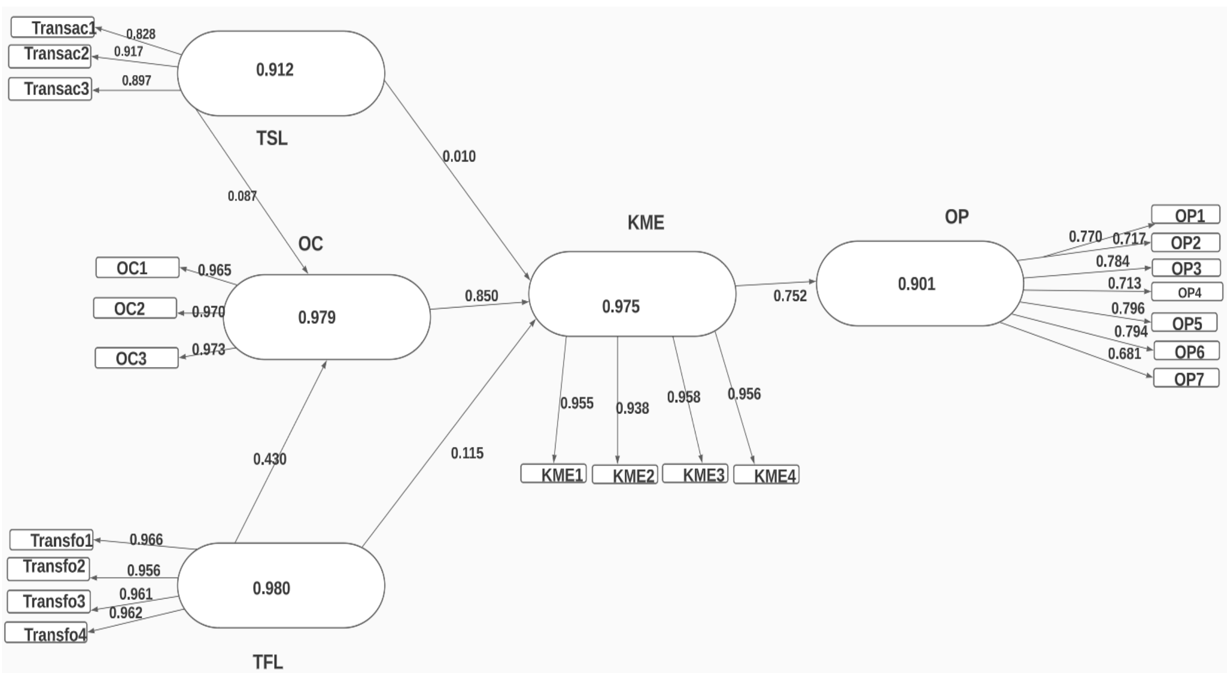


Figure 2 Measurement Model Results

Measurement Model Testing

Table 3 summarizes the quality metrics for the measurement model, including tests for convergent validity, internal consistency reliability, and discriminant validity with a significance level (alpha) of 0.05 and bias-corrected (BC) for interval analysis. All loadings exceed the recommended value of 0.70, except for OP7(removed from the model), accounting for a minimum of 50% of the variance for related constructs. AVE levels for all latent variables were higher than the recommended minimum value of 0.50 (Hair et al., 2019), affirming convergent validity. Internal consistency was compared with accepted values of 0.60 and 0.70 for Cronbach's alpha and composite reliability (CR), respectively. All constructs passed, indicating no issues with internal consistency (Hair et al., 2019). Two criteria of Heterotrait-monotrait (HTMT) were reviewed to assess discriminant validity, including an HTMT ratio of correlations score below the cutoff of 0.85

(Hair et al., 2020) and an HTMT interval not containing one, considering a confidence level (CI) of 95% (Hair et al., 2019; Wong, 2019). All constructs passed both criteria, confirming discriminant validity.

Table 3 Measurement Model Test Summary (loading >0.7)

Latent Variables	Indicators	Loadings	Convergent Validity		Internal Consistency Reliability		Discriminant Validity	
			rho_a	AVE	α	Composite Reliability	HTMT	
		>0.70	>0.50	>0.50	>0.60	>0.70	<0.85?	HTMT
OP	OP1	0.781	0.863	0.593	0.862	0.897	Yes	Yes
	OP2	0.742						
	OP3	0.784						
	OP4	0.710						
	OP5	0.793						
	OP6	0.803						
KME	KME1	0.955	0.966	0.906	0.965	0.965	No	Yes
	KME2	0.938						
	KME3	0.958						
	KME4	0.956						
OC	OC1	0.965	0.968	0.939	0.968	0.979	Yes	Yes
	OC2	0.970						
	OC3	0.973						
TSL	Transact1	0.828	0.855	0.777	0.855	0.912	Yes	Yes
	Transact2	0.917						
	Transact3	0.897						
TFL	Transfo1	0.966	0.972	0.973	0.972	0.980	Yes	Yes
	Transfo2	0.956						
	Transfo3	0.961						
	Transfo4	0.962						

Structural Model Testing

In this research, PLS-SEM modeling, using SmartPLS, was applied to discern cause and effect relationships of the structural model. The model was estimated utilizing the PLS algorithm with complete bootstrapping using the bias-corrected confidence interval method, two-tailed test, 5,000 subsamples, and mean replacement for missing values. Figure 3 depicts the structural model's hypothesized research model, while Table 4 summarizes the model estimation results using bootstrapping. R-squared is a goodness-of-fit measure for linear regression models. We observed 84% of the variance in the KME, 25% in the OC, and 56% for OP.

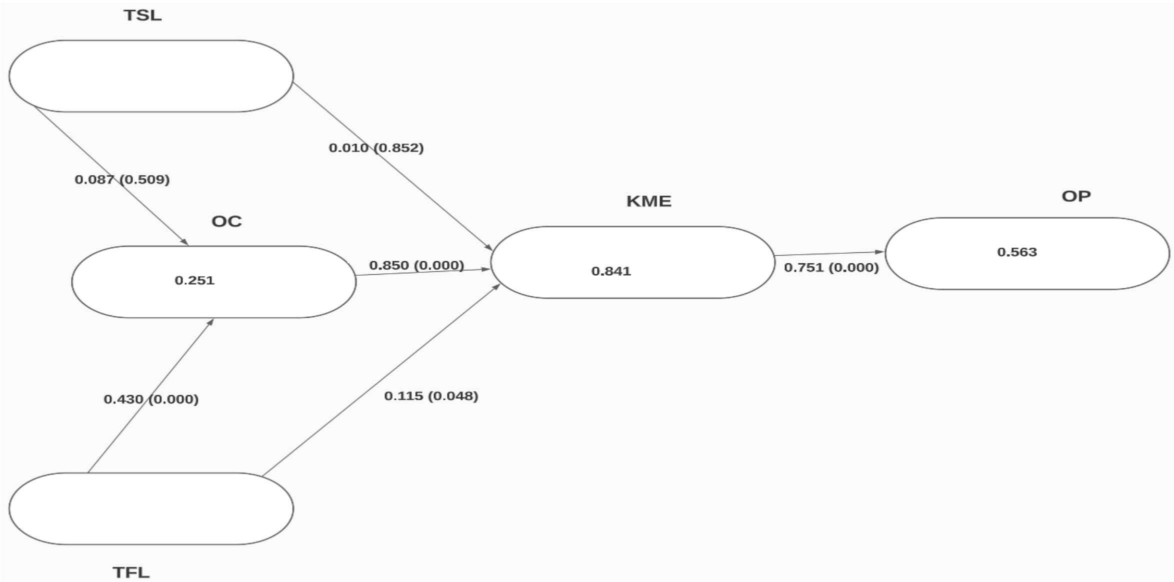


Figure 3 Structural Model Results with Endogenous Latent Variables

Table 4 Endogenous Variable Summary

Endogenous Latent Variables	Coefficient of Determination (R2)			R2 Adjusted			
	Value	Confidence Interval (BC)	p-value	Value	Confidence Interval (BC)	p-value	Predictive Relevance Q2
KME	0.841	[0.786, 0.882]	0.000	0.838	[0.782, 0.879]	0.000	0.756
OC	0.251	[0.107, 0.391]	0.001	0.24	[0.094, 0.381]	0.002	0.223
OP	0.563	[0.450, 0.657]	0.000	0.56	[0.445, 0.654]	0.000	0.325

Blindfolding, utilizing an Omission Distance (OD) of 7, was the technique used to calculate Stone-Geisser's Q2, which determines the predictive relevance of the model (Wong, 2019). All three endogenous latent variables have a Q2 greater than zero, which shows that the model has predictive relevance. In addition, we observed better than moderate predictive relevance for OC and OP and a substantial predictive relevance for KME.

Table 5 summarizes examining exogenous variables' effects on corresponding endogenous variables. This includes the mediating effects of OC. Considering VIF values of less than 5.0, We observed that multicollinearity is not a concern for all hypothesized relationships (Hair et al., 2013). Based on path coefficients and significance, there is support for hypotheses H4, H3, and H8 at an alpha of 0.01 (***) and H1 at an alpha of 0.05 (**). H2 and H5 were not significant.

Table 5 Structural Model Test Summary

Hypothesis	Relationship	VIF <5.0	Path Coefficient	Confidence Interval (BC)	p-value	Effect Size (f2)	Results
H1	TFL -> KME	2.774	0.115	[0.002, 0.227]	0.048**	0.03	Supported
H2	TSL -> KME	2.537	0.01	[-0.096, 0.116]	0.852	0.000	Not supported
H3	TFL -> OC	2.527	0.43	[0.192, 0.657]	0.000	0.098	Supported
H4	TSL -> OC	2.527	0.087	[-0.183, 0.324]	0.509	0.004	Not supported
H5	OC->KME	1.335	0.85	[0.783,0.898]	0.000***	3.405	Supported
H6	KME -> OP	1.000	0.75	[0.67, 0.81]	0.000***	1.29	Supported

Note. ***p<0.001, **p<0.01, *p<0.05.

DISCUSSION

Overall, the results suggest that transformational leadership and organizational culture directly influence the efficacy of KM in organizations in the context of HEI. As a result, KM has a positive influence on the performance in the context of HEI. The research also supports the hypothesis that organizational culture mediates the effects of transformational leadership on KM effectiveness. However, the mediating role of organizational culture in the relationship between transactional leadership on KM was not significant. Furthermore, transactional leadership did not significantly influence KM and organizational culture. The findings of this study indicate that transformational leaders are more pertinent to the implementation of knowledge management than transactional leaders. One possible explanation is that leaders that embrace transactional leadership styles tend to follow a more structured approach to employee relations. This type of leader does not allow for innovations. They do not focus on working conditions, such as an employee's connection with their executive, which can drive engagement and satisfaction. Transactional leaders who reward or acknowledge agreed-upon performance objectives without intellectual stimulation or consideration of subordinates' individual needs are not likely to attract, invigorate or retain employees (Avolio & Bass, 1995). Such leadership style is not particularly compatible in the context of HEI where shared governance, innovation and creativity are prominent and valued. The pressure of present-day competitive environment further highlights the need for leaders need to adjust to these changes through transformational leadership in such context.

Regarding RQ1, transformational leadership was found to affect KM processes significantly and positively in the context of HEI (H1 is supported). This is consistent with prior studies where transformational leadership has been demonstrated to positively influence KM effectiveness (Ather & Awan, 2021; Brandt et al., 2016; Sayyadi, 2019; Ugwu & Okore, 2020). Transformational leaders can cope with changes by proactively engaging with conflicting views in formulating a compelling vision for the future. As HEI are experiencing rapid changes due to technological advances such as Massive Open Online Courses (MOOCs) a transformational leader can promote KM values to motivate, inspire, and encourage employees to be involved in KM activities. However, the study found no significant relationship between transactional leadership and KM effectiveness (H2 is not supported). These results differs from the findings of Ather and Awan (2021) and Ugwu and Okore (2020) where both transformational and transactional leadership were found to have a significant relationship with knowledge management practices, supports the findings by Crawford (2005) who found no significant relationship between transactional leadership and knowledge management. This can be explained by the nature of transactional leadership were there is less tendency to focus on the personal development of the followers and more centered on goal attainment (Crawford, 2005; Ugwu & Okore, 2020). Such approach is may not be compatible in a HEI context with its emphasis on shared governance, creativity and innovation, and relative autonomy of faculty.

Regarding RQ2, the results demonstrated that effective implementation of KM processes is significantly and positively influential to the performance within HEI (H6 is supported). This finding reveals that effective KM processes (acquisition, sharing, and utilization) within HEI can increase research productivity, student satisfaction, curriculum development, research ranking, undergraduate student success, student placement rates, and responsiveness to environmental challenges. This result is in line with the findings of (Ahmad et al., 2015). They performed a study in a university context and concluded a significant positive and direct relationship between KM processes and OP.

Regarding RQ3. Organizational culture was found as a mediator of the relationship between transformational leadership and knowledge management. These results confirm the findings by Elshanti (2017) that organizational culture fully mediates the relationship between transactional leadership and organizational learning. These results indicate that transformational leadership is positively related to organizational culture, therefore provide support for H3. In essence, transformational leaders support the mission, involvement, consistency, and adaptability of their followers within their organizations. These leaders can change their culture by realigning the organization’s culture with a new vision and changing its shared assumptions, values, and norms (Bass & Avolio, 1994). These results also showed that organizational culture (adaptability, consistency, involvement, and mission) has a positive effect on knowledge management effectiveness, providing support for H5. These result mean that organizational culture has a positive and significant effect on KM processes. The findings draw attention to creating an organizational culture that encourages KM. The finding signifies the relative importance of culture. This may be because culture determines the fundamental beliefs, values, and norms regarding the why and how of knowledge generation, sharing, and utilization in an organization (Aldulaimi, 2015) and that culture capture the role of leadership as a mediator. This finding reinforces the necessity for attention for creating an organizational culture that encourages KM. The research also examined the relationship between transactional leadership and organizational culture and found no significant relationship, i.e., H4 is not supported. In the current complex knowledge environment, HEI must constantly innovate, analyze, evolve, investigate, respond to, and predict threats and opportunities. Therefore, they need transformational leaders that can innovate, encourage, and motivate followers as opposed to transactional leaders tend to follow a more structured approach to employee relations.

Although the study presented insights into the role of leadership and culture in supporting KM processes as well as the relationship between KM and organizational performance in the context of HEI, several limitations open up new opportunities for future research. One of the limitations of the study is that this study was interested in finding whether organizational culture plays a mediating role in the relationship between leadership style and knowledge management. Antonakis (2001) pointed out that leadership can shape the cultural traits of their organization's culture. However, organizational culture influences leaders' actions. Future research could be undertaken to establish the causality of this relationship and how this relationship develops. Another limitation is that in this study data were gathered using a self-reporting web survey instrument. This data collection technique could bias responses by members who report their perceptions of leadership, culture, knowledge management, and organizational performance. In this research, faculty were overly represented. While faculty are a significant consistent, future research could further explore sensitivity of the results with a focus on other constituents, e.g., administration and staff. Finally, a cultural issue may be observed in this study. The study was entirely conducted in a US context. Institutions of higher education have different cultures and structures, and possibly sub-cultures within institutions along disciplinary lines. Consequently, the finding is limited to the universities and colleges in the United States and may need to be more generalizable to other developed countries. Further studies could be conducted in other developed countries to address this limitation. Future researchers might also explore the role of organizational culture as a moderator of the relationship between knowledge management and leadership style.

CONCLUSION

This study examined the relationships among leadership style, organizational culture, KM effectiveness, and organizational performance in the context HEI in the United States. Overall results suggested that transformational leadership supports knowledge management processes, with organizational culture acting as a mediator. These findings identify transformational leadership as necessary in supporting culture and KM effectiveness. In addition, organizational culture was also found as a critical driver for KM effectiveness. Transactional leadership style was not related to knowledge management processes. This research answers the call by (Iqbal et al., 2019) who recommended that future studies consider the relationship between KM and organizational performance. They explained that such a study with different cultures and structures would add significant value to this field of study. While KM has been extensively studied, no study, to the best of our knowledge, has attempted to empirically combine all five constructs (KM, transformational leadership, transactional leadership, organizational culture, and organizational performance), into one single model for analysis.

This study could help HEI understand the type of leaders they need to promote a welcoming work environment and build up and sustain the knowledge management practice in their organizations. Specifically, it helps HEI identify which leadership styles is more likely to succeed in such context. This study informs HEI regarding the role of culture and leadership styles on KME. Increased knowledge about organizational culture can provide leaders, managers, and researchers unique insight into managing organizational culture to ensure that it aligns with the organization's ecosystem (Elshanti, 2017). In essence, and in the quest to maximize the benefits from knowledge management processes, leaders in a HEI context need to align with a transformational leadership style and seek to catalyze and culture supporting such processes.

APPENDIX A: SURVEY INSTRUMENT

Multifactor leadership questionnaire (MLQ), modified from Bass and Avolio (1994)

Leader form

Please answer all items on this answer sheet. If an item is irrelevant, or if you are unsure or do not know the answer, leave the answer blank. Twenty-eight descriptive statements are listed on the following pages. Judge how frequently each statement fits you. The word “others” may mean your peers, clients, direct reports, supervisors, and/or all these individuals.

Five-point Likert scale

Transactional Leadership

1. I provide others with assistance in exchange for their efforts
2. I re-examine critical assumptions to question whether they are appropriate
3. I focus attention on irregularities, mistakes, exceptions, and deviations from standards
4. I talk about my most important values and beliefs
5. I seek differing perspectives when solving problems
6. I talk optimistically about the future

7. I instill pride in others for being associated with me
8. I discuss in specific terms who is responsible for achieving performance targets

Transformational Leadership

1. I talk enthusiastically about what needs to be accomplished
2. I specify the importance of having a strong sense of purpose
3. I spend time teaching and coaching
4. I make clear what one can expect to receive when performance goals are achieved
5. I go beyond self-interest for the good of the group
6. I treat others as individuals rather than just as a member of a group
7. I act in ways that build others' respect for me
8. I concentrate my full attention on dealing with mistakes, complaints, and failures
9. I consider the moral and ethical consequences of decisions
10. I keep track of all mistakes
11. I display a sense of power and confidence
12. I articulate a compelling vision of the future
13. I direct my attention toward failures to meet standards
14. I consider an individual as having different needs, abilities, and aspirations from others
15. I get others to look at problems from many different angles
16. I help others to develop their strengths
17. I suggest new ways of looking at how to complete assignments
18. I emphasize the importance of having a collective sense of mission
19. I express satisfaction when others meet expectations
20. I express confidence that goals will be achieved

Rater form

This section is intended to describe your Leadership from your perspective. If you do not know the answer, leave it blank. For the following constructs, rate your immediate Manager regarding the elements listed from 1 to 5.

Transactional Leadership

1. Provides me with assistance in exchange for my efforts
2. Re-examines critical assumptions to question whether they are appropriate.
3. Focuses attention on irregularities, mistakes, exceptions, and deviations from standards
4. Talks about their most important values and beliefs
5. Seeks differing perspectives when solving problems.
6. Talks optimistically about the future
7. Instills pride in me for being associated with him/her
8. Discusses in specific terms who is responsible for achieving performance targets

Transformational Leadership

1. Talks enthusiastically about what needs to be accomplished
2. Specifies the importance of having a strong sense of purpose
3. Spends time teaching and coaching
4. Makes clear what one can expect to receive when performance goals are achieved
5. Goes beyond self-interest for the good of the group
6. Treats me as an individual rather than just as a member of a group

7. Acts in ways that builds my respect
8. Concentrates his/her full attention on dealing with mistakes, complaints, and failures
9. Considers the moral and ethical consequences of decisions
10. Keeps track of all mistakes
11. Displays a sense of power and confidence
12. Articulates a compelling vision of the future
13. Directs my attention toward failures to meet standards
14. Considers me as having different needs, abilities, and aspirations from others
15. Gets me to look at problems from many different angles
16. Helps me to develop my strengths
17. Suggests new ways of looking at how to complete assignments
18. Emphasizes the importance of having a collective sense of mission
19. Expresses satisfaction when I meet expectations
20. Expresses confidence that goals will be achieved

Knowledge management effectiveness, modified from Ahmad et al (2015)

For the following constructs, rate your school regarding the elements listed from 1 to 5.
Knowledge Acquisition

1. We hire new faculty and staff as a source for acquiring new knowledge.
2. We provide an open environment to our faculty and staff acquire new knowledge.
3. We actively observe and adopt the best practice in our higher education sector.
4. We continually gather information that is relevant to our operations and activities.
5. We list and define the knowledge we possess as well as any unavailable knowledge.
6. We obtain knowledge from different sources: students, partners, and faculty and staff.

Knowledge Sharing

1. We share information and knowledge necessary for the tasks.
2. We exchange knowledge between faculty and staff to achieve our goals with little time and effort
3. We developed information systems, like intranet and electronic bulletin boards to share information and knowledge
4. We promote sharing of information and knowledge between team members and the various departments.
5. Knowledge is shared between supervisors and subordinates.

Knowledge utilization

1. The school relies on established best-practices in guiding operations and making decisions.
2. Workflow diagrams are required and used in performing tasks.
3. The school effectively manages different sources and types of knowledge.
4. The school utilizes available knowledge in improving services provided to its students.
5. The school applies available knowledge to improve its performance.

Organizational culture, adapted from Denison and Mishra (1995); and Fey and Denison (2003)

For the following constructs, rate your school regarding the elements listed from 1 to 5.
Involvement

1. Our school continuously invest in the skills of faculty and staff.
2. Decisions are usually made at the level where the best information is available.
3. Working in your school is like being part of a team.

Consistency

1. There is a clear and consistent set of values in this school that governs the way you do business.
2. It is easy to reach consensus, even on different issues.
3. People from different departments still share a common perspective.

Adaptability

1. This school continually adopts new and improved ways to do work.
2. Student input directly influences our decision in this school
3. This school encourages and rewards those who take risk.

Mission

1. This school has a clear strategy for the future.
2. Leadership has clearly stated the objective we are trying to meet.
3. Our vision creates excitement and motivation for our faculty and staff.

Organizational performance (OP), Modified from Ahmad et al. (2015)

This section is intended to describe your organizational performance from your perspective. If you do not know the answer, leave it blank. For the following constructs, rate your school regarding the elements listed from 1 to 5.

1. Students satisfaction of our school is better as compared to peer schools.
2. Curriculum development process of our school is better as compared to peer schools.
3. Responsiveness of our school is better as compared to peer schools.
4. Research productivity of our school is better as compared to peer schools.
5. Research ranking of our school is better as compared to peer schools.
6. First-year undergraduate student success rate is better as compared to our peers.
7. Placement rates of our schools is better as compared to our peer schools.

Open-ended questions

1. In your opinion, which factors/KM processes are the most effective in implementing knowledge management practices at your school?

2. In your judgment, what are the primary barriers to effective knowledge management initiatives among administration, staff, and faculty at your school?

Demographic

1. What is your Gender? ____Male____Female____Other

2. What is your Age?

____Under 31 years? ___31 to 35 years ___36 to 40 years
____41 to 45 years____46 to 50 years___over 50 years

3. What is your position?

___Faculty member ___Faculty Chair ____Assistant Director

___Director ___Associate Dean ____Associate Vice-President
___Dean ___Associate Provost ____Manager ___Senior Manager ____Other

4. Do you have working knowledge of Information Systems or data management? (Check all that apply)

___ Working knowledge ___ Expert ___ Familiar ___ Novice

5. How long have you worked in your current area/position?

___ Less than 1 year ___ More than 1 and less than 2 ___ More than 2 and less than 3
___ More than 3 and Less than 4 ___ More than 4 and less than 5 ___ More than 5 years

6. How long (in years) have you worked for the school?

___ Less than 1 year ___ More than 1 and less than 2 ___ More than 2 and less than 3
___ More than 3 and Less than 4 ___ More than 4 and less than 5 ___ More than 5 years

7. Please indicate the option that best describes your school:

______Public ____Private, not-for-profit

8. Please indicate the level(s) of degrees awarded by your school (select all that apply):

*Note – this is your degree, but the degrees offered at your school
____ Certificate ____ Associate __ Bachelor____Masters___Doctoral ___Professional (e.g., MD, JD, DDS)

9. Please indicate your school's name (the school names will remain anonymous):

*Note – the school's name will remain anonymous

REFERENCES

Acar, A. Z. (2012). Organizational Culture, Leadership Styles and Organizational Commitment in Turkish Logistics Industry. Social and Behavioral Sciences, 58(12), 217–226. https://doi.org/10.1016/j.sbspro.2012.09.995

Ahmad, M. I., Ahmad, A., & Riaz, R. (2020). Utilization of Knowledge Management Process (KMP) Model to Predict Job Performance in Higher Education Sector of Pakistan. Paradigms, 14(1), 148–152.

Ahmad, N., Lodhi, M. S., Zaman, K., & Naseem, I. (2015). Knowledge Management: A Gateway for Organizational Performance. Journal of the Knowledge Economy, 8(3), 859–876. https://doi.org/10.1007/s13132-015-0282-3

Ahmady, G. A., Nikooravesh, A., & Mehrpour, M. (2016). Effect of Organizational Culture on Knowledge Management Based on Denison Model. Procedia - Social and Behavioral Sciences, 230, 387–395. https://doi.org/10.1016/j.sbspro.2016.09.049

Akhavan, P., Zahedi, M. R., Dastyari, A., & Abasaltian, A. (2014). The Effect of Organizational Culture and Leadership Style on Knowledge Management in Selected Research Organizations. International Journal of Scientific Management and Development, 2(9), 432–440.

Aldulaimi, S. H. (2015). Exploring the Effect of Organizational Culture, Leadership and Strategy on Organizational Effectiveness with Mediating Effect of Knowledge Management. International Journal of Economics, Commerce and Management, 3(4), 20.

Al-Qarioti, M. Q. A. (2015). The Impact of Knowledge Management on Organizational Performance: An Empirical Study of Kuwait University. Eurasian Journal of Business and Management, 3(4), 36–54. https://doi.org/10.15604/ejbm.2015.03.04.004

Alshahrani, A. (2018). Critical Success Factors of Knowledge Management in Higher Education Institutions: A Comparative Study Between Western Sydney University in Australia and King Fahd Security College in Saudi Arabia. [Thesis]. Western Sydney University.

Antonakis, J. (2001). The Validity of the Transformational, Transactional, and Laissez-Faire Leadership Model as Measured by the Multifactor Leadership Questionnaire (MLQ 5X). [Thesis]. Walden University.
Ather, T., & Awan, R.-N. (2021). Effect of Knowledge Management Practices and Leadership Styles of Heads of Departments on University Teachers' Performance. Bulletin of Education and Research, 43(159), 181–200.

Avolio, B. J., & Bass, B. M. (1995). Individual Consideration Viewed at Multiple Levels of Analysis: A Multi-Level Framework for Examining the Diffusion of Transformational Leadership. The Leadership Quarterly, 6(2), 199–218. https://doi.org/10.1016/1048-9843(95)90035-7

Avolio, B. J., Bass, B. M., & Jung, D. I. (1999). Re-examining The Components of Transformational and Transactional Leadership Using the Multifactor Leadership Questionnaire. Journal of Occupational and Organizational Psychology, 72(4), 441–462. https://doi.org/10.1348/096317999166789

Baker, J., Parasuraman, A., Grewal, D., & Voss, G. B. (2002). The Influence of Multiple Store Environment Cues on Perceived Merchandise Value and Patronage Intentions. Journal of Marketing, 66(2), 120–141. https://doi.org/10.1509/jmkg.66.2.120.18470

Bass, B. M., & Avolio, B. J. (1994). Transformational Leadership and Organizational Culture. International Journal of Public Administration, 17(3–4), 541–554. https://doi.org/10.1080/01900699408524907

Brandt, T., Laitinen, E. K., & Laitinen, T. (2016). The Effect of Transformational Leadership on the Profitability of Finnish Firms. International Journal of Organizational Analysis, 24(1), 81–106. https://doi.org/10.1108/IJOA-03-2014-0744

Chen, J., & Nonaka, I. (2022). The Routledge Companion to Knowledge Management (1st ed.). Routledge. https://doi.org/10.4324/9781003112150

Chiu, C.-N., & Chen, H.-H. (2016). The Study of Knowledge Management Capability and Organizational Effectiveness in Taiwanese Public Utility: The Mediator Role of

Organizational Commitment. SpringerPlus, 5(1520). https://doi.org/10.1186/s40064-016-3173-6

Crawford, C. B. (2005). Effects of Transformational Leadership and Organizational Position on Knowledge Management. Journal of Knowledge Management, 9(6), 6–16. https://doi.org/10.1108/13673270510629927

Deichmann, D., & Stam, D. (2015). Leveraging Transformational and Transactional Leadership to Cultivate the Generation of Organization-Focused Ideas. The Leadership Quarterly, 26(2), 204–219. https://doi.org/10.1016/j.leaqua.2014.10.004

Denison, D. R., & Mishra, A. K. (1995). Toward a Theory of Organizational Culture and Effectiveness. Organization Science, 6(2), 204–223. https://doi.org/10.1287/orsc.6.2.204

Dhamdhere, S. N. (2015). Knowledge Management Strategies and Process in Traditional Colleges: A Study. International Journal of Information Library and Society, 4(1), 34–42.

Elshanti, M. (2017). Transformational Leadership Style and Organizational Learning: The Mediate Effect of Organizational Culture. International Journal of Economics & Management Sciences, 15(02), 1–14. https://doi.org/10.4172/2162-6359.1000483

Fernández-López, S., Rodeiro-Pazos, D., Calvo, N., & Rodríguez-Gulías, M. J. (2018). The Effect of Strategic Knowledge Management on the Universities' Performance: An Empirical Approach. Journal of Knowledge Management, 22(3), 567–586. https://doi.org/10.1108/JKM-08-2017-0376

Fey, C. F., & Denison, D. R. (2003). Organizational Culture and Effectiveness: Can American Theory be Applied in Russia? Informs, 14(6), 686–706.

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing Measurement Model Quality in PLS-SEM Using Confirmatory Composite Analysis. Journal of Business Research, 109, 101–110. https://doi.org/10.1016/j.jbusres.2019.11.069

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial Least Squares Structural Equation Modeling: Rigorous Applications, Better Results and Higher Acceptance. Long Range Planning, 46(1), 1–12. https://doi.org/10.1016/j.lrp.2013.01.001

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to Use and How to Report the Results of PLS-SEM. European Business Review, 31(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Iqbal, A., Latif, F., Marimon, F., Sahibzada, U. F., & Hussain, S. (2019). From Knowledge Management to Organizational Performance: Modelling the Mediating Role of Innovation and Intellectual Capital in Higher Education. Journal of Enterprise Information Management, 32(1), 36–59. https://doi.org/10.1108/JEIM-04-2018-0083

Khakpour, A. (2015). Knowledge Management in Educational Organizations: Opportunities and Challenges. 7th International Knowledge Management Conference.

Kidwell, J. J., Linde, K. M. V., & Johnson, S. L. (2000). Applying Corporate Knowledge Management Practices in Higher Education. Educause Quarterly, 23(4), 28–33.

Kishton, J. M., & Widman, K. F. (1994). Unidimensional Versus Domain Representative Parceling of Questionnaire Items: An Empirical Example. Educational and Psychological Measurement., 54(3), 757–765.

Lee, A. (2018). Differences in Leadership Behaviors of Principals and Success in the School Turnaround Processes. [Thesis]. Louisiana Tech University.

Millar, C., Peters, K., & Millar, H. (2018). Culture, the Missing Link in Value Creation and Governance in Knowledge-Intensive Institutions. John Wiley & Sons, Ltd., 18(1). https://doi.org/10.1002/pa.1702

Nam Nguyen, H., & Mohamed, S. (2011). Leadership Behaviors, Organizational Culture and Knowledge Management Practices. Journal of Management Development, 30(2), 206–221. https://doi.org/10.1108/02621711111105786

Nawaz, D. M. N., Phil, M., Gomes, D. A. M., Com, M., & Ed, B. (2014). Review of Knowledge Management in Higher Education Institutions. European Journal of Business and Management, 6(7), 9.

Quarchioni, S., Paternostro, S., & Trovarelli, F. (2022). Knowledge Management in Higher Education: A Literature Review and Further Research Avenues. Knowledge Management Research & Practice, 20(2), 304–319. https://doi.org/10.1080/14778238.2020.1730717
Rahman, M. (n.d.). Approaches to Managing Costs in American Higher Education.
Razmerita, L., Kirchner, K., & Nielsen, P. (2016). What Factors Influence Knowledge Sharing in Organizations? A Social Dilemma Perspective of Social Media Communication. Journal of Knowledge Management, 20(6), 1225–1246. https://doi.org/10.1108/JKM-03-2016-0112

Rodríguez-Gómez, D., & Gairín, J. (2015). Unravelling Knowledge Creation and Management in Educational Organisations: Barriers and Enablers. Knowledge Management Research & Practice, 13(2), 149–159. https://doi.org/10.1057/kmrp.2013.37

Rowley, J. (2000). Is Higher Education Ready for Knowledge Management? International Journal of Educational Management, 14(7), 325–333. https://doi.org/10.1108/09513540010378978

Sahibzada, U. F., Cai, J., Latif, K. F., & Sahibzada, H. F. (2020). Knowledge Management Processes, Knowledge Worker Satisfaction, and Organizational Performance: Symmetric and Asymmetrical Analysis. Aslib Journal of Information Management, 72(1), 112–129. https://doi.org/10.1108/AJIM-10-2019-0276

Sayyadi, M. (2019). How Effective Leadership of Knowledge Management Impacts Organizational Performance. Business Information Review, 36(1), 30–38. https://doi.org/10.1177/0266382119829643

Seyedyousefi, N., Hosseini Fard, S. M., & Tohidi, F. (2016). The Role of Organizational Culture in Knowledge Management. Mediterranean Journal of Social Sciences, 7(5). https://doi.org/10.5901/mjss.2016.v7n5p412

Shirini, N., & Xenikou, A. (2022). Investigating the Mediating Role of Organizational Culture in the Relationship Between Transformational Leadership and Organizational Effectiveness. International Journal of Innovation in Management Economics and Social Science, 2(1), 85–101.

Si Xue, C. T. (2017). A Literature Review on Knowledge Management in Organizations. Research in Business and Management, 4(1). https://doi.org/10.5296/rbm.v4i1.10786

Sufyan Areed, Said A. Salloum, & Khaled Shaalan. (2021). The Role of Knowledge Management Processes for Enhancing and Supporting Innovative Organizations: A Systematic Review. In Recent Advances in Intelligent Systems and Smart Applications (Vol. 295, pp. 143–161). Springer International Publishing. https://doi.org/10.1007/978-3-030-47411-9_8

Sunalai, S., & Beyerlein, M. (2015). Exploring Knowledge Management in Higher Education Institutions: Processes, Influences, and Outcomes. Academy of Educational Leadership Journal, 19(3), 289–308.

Tierney, W. G., & Lechuga, V. M. (2004). Restructuring. Shared Governance in Higher Education. Wiley Periodicals, Inc., 127.

Ugwu, C. I., & Okore, A. M. (2020). Transformational and Transactional Leadership Influence on Knowledge Management Activities of Librarians in University Libraries in Nigeria. Journal of Librarianship and Information Science, 52(3), 864–879. https://doi.org/10.1177/0961000619880229

Veer Ramjeawon, P., & Rowley, J. (2017). Knowledge Management in Higher Education Institutions: Enablers and Barriers in Mauritius. The Learning Organization, 24(5), 366–377. https://doi.org/10.1108/TLO-03-2017-0030

Wiley University Services. (2021, November). Top Challenges Facing U.S. Higher Education. [Wiley University Services]. Top Challenges Facing U.S. Higher Education. https://universityservices.wiley.com/top-higher-education-challenges/

Wong, K.-K. (2019). Mastering Partial Least Squares Structural Equation Modeling (PLS-SEM) with SmartPLS in 38 Hours. iUniverse.

DECISION SCIENCES INSTITUTE

The Standardized Development Strategy of Shanghai Community Sports Volunteer Service

Wu Xinyu
Tongji University
Email: wuxinyubsu@126.com

ABSTRACT

In order to promote the development of public sports volunteer service in Shanghai, this study aims to propose effective strategies for standardizing Shanghai's public sports volunteer service. An analysis of the current state of Shanghai's sports volunteer service reveals several issues, including a weak volunteer service atmosphere, inadequate social mobilization, uncoordinated organizational management, obstructed information channels, and a lack of targeted activities. In response, this article suggests that the standardization of Shanghai's sports volunteer service can be enhanced by implementing mechanisms for social mobilization, organizational coordination, training management, project promotion, incentives, and evaluation.

KEYWORDS: Sports volunteer service, Social governance, Standardization, Development strategies

INTRODUCTION

In line with the Shanghai Urban Master Plan (2017-2035) and the Shanghai Global Famous Sports City Construction Outline, clear strategic goals have been set for Shanghai's urban construction and sports development. By 2025, Shanghai aims to become a globally renowned sports city, world-class international sports event capital, an important sports resource allocation center at home and abroad, and a vibrant sports technology innovation platform. By 2035, Shanghai aims to become a global city of excellence, innovation, humanity, ecology, and a modern international socialist metropolis with global influence. To achieve these targets, targeted planning has been made for nationwide fitness activities in Shanghai, including the distribution of community citizen fitness activity centers throughout each community. Additionally, innovative urban governance methods must be implemented, refined management practices strengthened, and a new paradigm of social governance that conforms to the characteristics and laws of mega cities created according to the "Reply of the State Council on the Overall Urban Planning of Shanghai."

It is evidently that the development of a standardized operation mode for the voluntary service system for national fitness in Shanghai is necessary to build a world-renowned sports city by 2025 and move towards an outstanding global city plan by 2035. Further standardization of the operation mode for the Shanghai National Fitness Volunteer Service System has become an essential issue that urgently needs to be studied as it guides and drives the collaborative construction of other mass sports undertakings.

LITERATURE REVIEW

In Western countries, research on "sport volunteering" is more common.

Firstly, individual case studies related to sport volunteering have been conducted. A research explored the relationship between social capital and sport volunteering in two Canadian communities, and suggested that controlling variables such as gender, language, and age could establish a closer relationship between social capital and sport volunteering （Harvey，2007). Another research investigated the object, function, and affairs of sports organizational systems, analyzed the characteristics of public sports volunteering, and put forward targeted strategic recommendations accordingly (Aesung, 2010). In addition, a study analyzed the European Union's sport volunteering system, believing that by incorporating EU volunteers' social work into practical activities through sports organizations, young people's sense of social responsibility could be enhanced (E. Esra, 2014).

Secondly, the relational perspective has been examined. A study pointed out that improving security and incentive mechanisms for sport volunteers as service providers can significantly increase recruitment and improve their job performance (Wicker, 2020). Another research focused on the relationship between public sport volunteering and education, suggesting that it has broad benefits for both university students and society, and outlining future development paths combining university student volunteering opportunities with sports development plans (Hayton, 2018). A related research particularly focused on how public sport volunteering and social support interact: investigating the advantages, disadvantages, and potential human resource management influences of involving disabled volunteers in sports volunteering (Kappelides，2019).

Thirdly, the efficiency improvement methods have been explored. A study found a significant correlation between participation motivation and satisfaction with sport volunteering (Hyunmi, 2004). Another research discovered that transferring small trusts managing public sports facilities to volunteers has economic benefits, meeting local leisure needs (Findlay-King, 2018). A group conducted an empirical analysis of the roles and subjective well-being of sport volunteers in 28 European countries, concluding that the diversity of roles undertaken in sport volunteering has a positive impact on individual happiness (Wicker, 2020).

In summary, related researches mainly focus on solving micro-level problems in practice, while paying less attention to macro-level studies, putting more emphasis on investigating the economic value of public sport volunteering. However, there has yet to be a systematic investigation or standardized scientific construction.

METHODS

Literature Review Method

Using the literature review method, summarize the relevant research results on the standardization of sports volunteer services and public sports services.

Case Study Method

The case study method was used to collect and integrate the advanced experiences of the standardized development of the public sports volunteer service system in the United States, Britain, Japan and Taiwan, China.

Expert Interviews

Using expert interviews to understand the current difficulties and future directions in the development of public sports volunteer service in China, providing a basis for policy research.

RESULTS

Case Analysis of Mature Community Sports Services

This study integrated relatively mature sports volunteer service in the United States, Britain, Japan, and Taiwan, China. A comparative analysis is presented horizontally and vertically across four parts: volunteer scale, organizational form, service content, and incentive methods. Based on comparative analysis, it can be concluded that there are some commonalities in the operating models of mature public sports volunteer service systems both domestically and internationally.

Large Scale of Volunteer Services

The precondition for standardizing the operation of the national fitness volunteer service system is the formation of a substantial volunteer community. There is a large number of experienced public sports volunteers both domestically and internationally, and a favorable volunteering atmosphere of "everyone aspires to be a volunteer" has been established. By promoting and facilitating volunteer service actions throughout society, we aim to foster a supportive environment for volunteering and expand the national fitness volunteer team accordingly.

The Voluntary Service System is Relatively Complete

The standardization of public sports volunteer service systems both domestically and abroad is reflected in various mandatory laws and regulations, as well as specific rules governing public sports volunteering. An integrated institutional system serves as a fundamental guide for the practice of public sports volunteer service. Based on the experiences of public volunteer service systems globally, adhering to legal frameworks is a necessary approach towards constructing a standardized national fitness volunteer service system in China.

The Management of Volunteer Service Organizations is More Scientific

The public sports volunteer service system must cater to the volunteer service demands that arise during the physical exercise routines of all residents. Effective information management and organizational coordination among public sports activities, volunteers, and residents is crucial to promoting the public sports volunteer service system in an orderly fashion. The scientific aspect of organizational management serves as the bedrock for the construction and standardization of the national fitness volunteer service system.

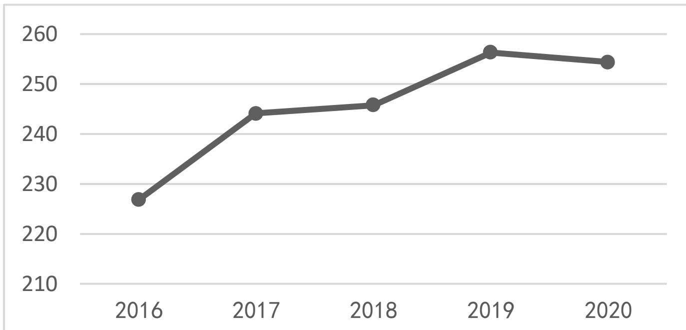
Voluntary Service Incentives and Guarantees are Relatively Reasonable

The engagement of volunteers in public sports volunteer services with the values of unity, friendliness, and dedication is praiseworthy, promoted, and endorsed by society. Funding basic insurance, food, and transportation ensures that passionate volunteers are supported and motivates them to enhance their commitment to the cause. This incentive and guarantee system for national fitness volunteers serves as a critical aspect in implementing the sustainable operational model of national fitness volunteer services.

Current Situation of Public Sports Development in Shanghai

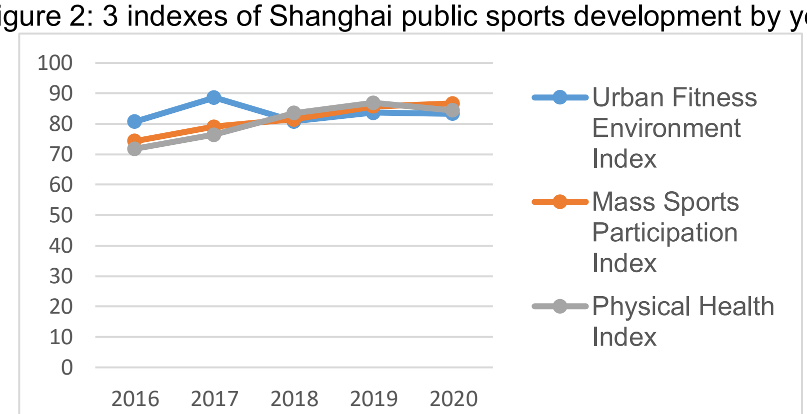
As one of China's leading super first-tier cities, Shanghai is committed to establishing itself as a globally renowned sports city. The development of sports cities involves not only the construction of physical infrastructure, event organization, promotion of local brands, and the integration of sports culture, but also aims to create an environment that promotes nationwide physical exercise, facilitating daily activities for everyone and transforming sports into a way of life.

Figure 1: “300 Index” of Shanghai public sports development by year



The Shanghai Public Sports Development Index, or "300 Index," has observed an increasing trend from 2016 to 2020. Drawing on the Shanghai National Fitness Development Report, this study examines the current state of fitness development in three major sectors, namely urban fitness environment, mass sports participation, and physical health. As illustrated in Figure 2, prior to 2018, the index of fitness environment in Shanghai was slightly higher than that of sports participation and physical health. However, since 2018, changes in all three sub-indices have been aligning, which indirectly confirms Shanghai's collaborative capability and its ability to advance national fitness goals while enhancing urban fitness environment, improving sports participation rates, and enhancing the overall physical health of residents.

Figure 2: 3 indexes of Shanghai public sports development by year



Current Situation of Urban Fitness Environment

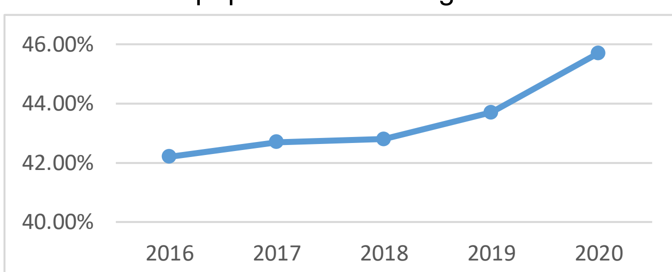
The development of urban fitness environments encompasses three key factors: government support, venue facilities construction, and citizen satisfaction. Thanks to the successful

implementation of the "13th Five Year Plan" for sports, Shanghai has achieved an unprecedented level of progress in the creation of its urban fitness environment. According to the 2020 Shanghai National Fitness Report, by the end of that year, the city had built a total of 1,954 kilometers of various types of citizen fitness trails (greenways) and cycling paths, alongside 17,556 citizen puzzle fitness parks, 2,714 citizen courts, 1,669 citizen fitness trails (greenways), 101 community citizen fitness centers, and 186 citizen gyms.
In 2020, Shanghai citizens expressed an 85.6% satisfaction rate towards sports venues, representing a continuous upward trend from 82% in 2015. The abovementioned statistics indicate that the city is continuously furthering the construction, renovation, and updating of the "Golden Corner and Silver Edge." By taking into account people's needs and transforming community sports facilities, Shanghai has gradually elevated and tailored local fitness infrastructure to meet the fitness requirements of its residents.

Current Situation of Participation in Mass Movements

From 2016 to 2020, the number of Shanghai citizens frequently engaging in physical exercise increased continuously, supported by the abundance and "proximity" of national fitness events and activities that provided opportunities for participation, thereby creating a "sports hot" and "fitness hot" atmosphere. In 2020, in response to the epidemic situation, Shanghai hosted its first online sports meet, which significantly promoted citizens' ability to enhance their physical and mental health levels and adjust their psychological state by utilizing an innovative fitness method integrating online and offline modes.
According to the 2020 Shanghai National Fitness Report, the third Shanghai Citizen Games held about 7,100 events and activities, including 800 online events and 6,300 offline events, with a total of 10.93 million participants.

Figure 3: The proportion of people who frequently participate in exercise to the permanent population in Shanghai



Current Situation of Physical Health of the Masses

The national fitness strategy advocates that the entire nation participate in physical fitness activities at least once a day, learn two or more fitness methods, and undergo an annual physical fitness test. To achieve this goal, all provinces and cities must actively promote national fitness guidance and monitor physical fitness, using scientific data indicators to assess nationwide physical fitness levels and fitness literacy. Public data revealed that in 2020, the physical fitness compliance rate of Shanghai citizens remained among the highest in the country, with relatively high rates of physical fitness excellence among adults and the elderly. However, the physical fitness excellence rate among the youth group only reached fifty percent. In terms of fitness guidance, Shanghai had a total of 62,086 social sports instructors by the end of 2020, accounting for 2.5% of the permanent population. There were 2.33 fixed fitness teams per thousand citizens, indicating considerable development of fitness guidance at scale.

Table 1: The Physical Fitness Compliance Rate of Shanghai Citizens in 2020		
GROUP	PASS RATE	EXCELLENCE RATE
Adults	98.9%	75.0%
Elderly	99.5%	72.9%
Youth	97.2%	50.6%

The Development Status of Community Sports Volunteer Service in Shanghai

The operation mechanism of the Shanghai National Fitness Volunteer Service is innovatively implemented through service delivery, primarily coordinated by the Shanghai Community Sports Association. Since 2015, the Association has commenced community sports services distribution throughout Shanghai to address personalized physical exercise requirements of residents. The service distribution encompasses fitness skills training, scientific fitness lectures, youth sports training, and public sports science lecture halls with progressive service forms and effective results. By establishing a three-level community sports service distribution mechanism at the city, district, and street level, along with a digital service network of "Internet plus + scientific fitness" to coordinate and integrate social resources, the Association has expanded the coverage of scientific fitness guidance benefiting more people through sports.

<u>Social Mobilization Mechanism</u>

Currently, there are four primary sources of public sports volunteers in Shanghai: those recommended by enterprises and institutions, sports teachers and university students, sports enthusiasts with professional skills, and doctors. However, the social sports instructors and sports enthusiasts are the mainstay in the Shanghai national fitness volunteer team based on the actual composition of personnel participating in the supply of national fitness volunteer services. Additionally, the majority of volunteers participating in the delivery of national fitness volunteer services are retirees. From the source channels of national fitness volunteers, it can be inferred that the present social mobilization mechanism of Shanghai's national fitness volunteer service mainly relies on cooperative units to strengthen the volunteer service team and provide certain class hour subsidies to distribution volunteers. The channels for social mobilization are relatively limited, lacking a comprehensive mobilization mechanism for the entire society. Overall, the existing social mobilization mechanism is relatively weak and has a low mobilizing effect.

<u>Organizational Coordination Mechanism</u>

The Shanghai Community Sports Association is the leading organization in promoting the operation of voluntary national fitness services in Shanghai. The Association was initiated at the end of 2013 after consultations with several communities, including Siping Community in Yangpu District, Xujiahui Community in Xuhui District, Xinjing Town Community in Changning District, and Yangjing Community's sports club in Pudong New Area. The association comprises five functional departments: the Competition Department, Distribution Department, Organization and Construction Department, Instructor Department, and Office. In addition, sports clubs are also uniformly organized and managed by region, resulting in a relatively perfect organizational and coordination mechanism. According to the official website of the Shanghai Community

Sports Association, the organization currently has 238 club members, including various community health clubs located on different districts, as well as a few enterprise health clubs. The primary tasks of the Association include managing relevant community public sports facilities; regularly organizing and carrying out community sports activities; providing scientific fitness guidance and sports skill training; developing and cultivating community sports volunteers; organizing the implementation of community sports service delivery; and promoting and popularizing community sports and fitness programs.

<u>Training Management Mechanism</u>

The professional development of the volunteer team is a crucial aspect of the Shanghai National Fitness Volunteer Service, and the associated volunteer training management mechanism continues to improve. However, at present, the training management of national fitness volunteers in Shanghai operates according to the principle of voluntariness, which may not ensure the effectiveness of volunteer training. Additionally, due to various limitations such as training time, location, content, and format, the existing training management fails to consider the actual circumstances of most volunteers. Consequently, many volunteers are unable to attend training sessions, leading to unsatisfactory results.

<u>Project Promotion Mechanism</u>

The promotion of Shanghai's national fitness volunteer service project involves not only collaborating with relevant city, district, and personnel but also engaging with social entities to collectively provide national fitness events and volunteer services. According to the "2020 Shanghai National Fitness Development Report," a total of 247 enterprises and social organizations partnered in Shanghai's national fitness events via public bidding in 2020. They organized over 6121 events and activities, with a overall participation of 8.7 million individuals, including 5814 in-person events with 4.81 million participants and 307 online competitions and activities with 3.99 million participants.

<u>Recognition and Incentive Mechanism</u>

The current volunteer commendation and incentive mechanism in Shanghai primarily relies on spiritual motivation, encouraging volunteers to uphold the values of "dedication, friendship, mutual assistance, and progress" in their service. While providing fitness guidance according to their own capabilities, the emphasis is less on material rewards. However, within the Shanghai National Fitness Volunteer Service Delivery Mechanism, instructors receive certain fees for classes and transportation expenses, while delivery personnel are given sports clothing. Outstanding delivery personnel at the end of the year are recognized, certified, and awarded bonuses. Despite these commendations, there exist constraints as some delivery personnel refuse to participate in non-delivery activities. To address such issues, the commendation incentive mechanism needs to be reinforced and enhanced.

<u>Evaluation and Inspection Mechanism</u>

The Shanghai Municipal Sports Bureau assesses the development of Shanghai's national fitness each year and publishes a report on its progress, utilizing data and research results to demonstrate its effectiveness. In evaluating the national fitness volunteer service system, it is imperative to use data to assess the number of distribution activities, volunteers, and citizens' physical and health levels. Additionally, evaluating the service quality of national fitness volunteers involves tracking volunteer hours, class hours, service recipients, feedback from local streets, and associations. Currently, the evaluation and inspection mechanism leans towards subjective evaluation to encourage volunteers to provide long-term and active services. However, a diversified evaluation mechanism should be established to enhance the effectiveness of volunteer services. This trend represents the future of fitness volunteer service system development.

The Current Problems of Community Sports Volunteer Service in Shanghai

The Atmosphere of Volunteer Service is Not Strong

The recruitment source for volunteers in Shanghai's national fitness volunteer service is limited, resulting in a lack of sustained effectiveness in the program. Despite possessing expertise and skills in sports, numerous community residents are reluctant to participate in the volunteer service due to time constraints and other factors, while insufficient enthusiasm weakens the overarching atmosphere of the program. Additionally, there exists a rather one-dimensional promotional strategy for advancing the values and goals of this volunteering initiative throughout the city. This approach fails to fully bridge the cognitive gap between promotion and the daily routines of residents, hampering efforts to truly cultivate a culture of volunteerism.

Insufficient Social Mobilization

The bulk of volunteers for the national fitness volunteer service program in Shanghai are largely middle-aged and elderly women; other groups such as sports researchers, university sports instructors and students, and community residents who possess sports-related expertise have yet to fully participate in the program. This represents untapped potential for bolstering the volunteer service corps, while mobilizing social support remains both an underlying prerequisite and cornerstone for fortifying Shanghai's national fitness volunteer team.

Inconsistent Organizational Management

Despite having a relatively advanced supply model for national fitness-related volunteer services, as a first-tier city in China, Shanghai has an enormous urban population with highly personalized and diverse exercise needs, which makes it difficult to satisfy the fitness requirements of each individual solely by relying on administrative powers to organize, manage, and coordinate a volunteer service system. The delivery of national fitness activities and the management of such volunteer services inherently depend on social forces. Therefore, integrating more, better-skilled, and more experienced third-party organizations into the organization and management of the national fitness volunteer service system is the cornerstone of advancing Shanghai's national fitness volunteer service.

Blocked Information Channels

There exist communication barriers between the information regarding physical activities and that of volunteer service in Shanghai's National Fitness Volunteer Service. Information related to the National Fitness Volunteer Service can only be accurately retrieved through relevant network channels under the Shanghai Community Sports Association. However, for ordinary citizens or those interested in joining the National Fitness Volunteer Service, searching for information channels is a relatively cumbersome process. The research team discovered, through online resource retrieval, that the Shanghai Volunteer Network has realized a highly intelligent and information-based design around volunteer recruitment, training, and information dissemination. Nevertheless, it was observed that no targeted recruitment of sports volunteers existed during the search for sports and sports volunteer activities. Evidently, volunteer needs during national fitness activities were not synchronously posted on the Shanghai Volunteer Website, and information channels weren't established. It can be inferred that Shanghai Volunteer Network, as a website that coordinates and manages volunteers in Shanghai, must comprehensively consider the types and service systems of volunteer services, including the National Fitness Volunteer Service system, to maximize the platform's value.

<u>Inaccurate Activity Implementation</u>

Upon analyzing data concerning the development of national fitness in Shanghai, it is apparent that an ironic situation exists: the age group with the most concerning physical health are teenagers, with an excellent rate lower than that of the elderly population. The underlying reason for this being that when undertaking volunteer activities for national fitness, organizers tend to prioritize the elderly group, who have more leisure time, and entrust the responsibility of improving the young people's physical health to school sports, thus neglecting this population's national fitness needs. Additionally, there is a lack of fitness activities catered to the youth and middle-aged population resulting from the limited implementation, precision, systematization, and completeness of national fitness activities. Such inadequacies make it difficult to discover and cultivate potential volunteers, who grew up in this environmental atmosphere. Therefore, implementing targeted national fitness activities tailored to different populations' physical health levels and personalized needs would be an effective way to fortify the national fitness volunteer team.

DISCUSSION AND CONCLUSIONS

Promoting the Standardization of Social Mobilization Mechanisms

The methods for effectively promoting the spirit of volunteer service and cultivating a volunteer-friendly environment involves two main aspects: internal affirmation and encouragement of volunteer organizations, as well as external publicity and promotion of volunteer organizations. The promotion and cultivation of volunteer service and the encouragement of more individuals to join the volunteer team is an ongoing issue in operating the national fitness volunteer service system. Internally, it is essential to strengthen the humanization of management. Personal responsible for volunteer management and relevant departments should consider the psychological state of volunteers, provide adequate emotional and material support, affirm and encourage their efforts, and increase confidence and patience in participating in national fitness volunteer services continuously. By doing so, retaining current and attracting future national fitness volunteers, while activating social mobilization mechanisms, can be achieved. Furthermore, to promote and popularize the spirit of volunteer service, diversification and effectiveness of promotional channels are crucial. In the era of short videos, people have access to various sources of information through personalized selections, making the combination of traditional media promotion methods with new media promotion paths necessary. Thus, expanding promotional channels is encouraged to reinforce the spirit of volunteer service in the new era.

Emphasizing the Standardization of Organizational Coordination Mechanisms

According to research, though there is a large number of registered social sports instructors in Shanghai, the actual participation rate in national fitness volunteer services is relatively small, resulting in a huge gap. To address this issue, it is important to lead the strengthening of resource integration among the existing sports social instructor teams in Shanghai, enhancing the mobility, sense of responsibility, and belongingness of the social sports instructor team. Additionally, collaborating with high-quality social forces to organize and manage volunteer services for national fitness is also crucial. Currently, numerous enterprises and organizations driving the supply of volunteer services have emerged in China's social field. They possess extensive experience and are committed to volunteer supply, training, and management in competitions and events. Collaborating with these social forces and integrating official and social organizations flexibly while reasonably utilizing social resources are vital measures to standardizing the national fitness volunteer service system. Another significant aspect is enhancing the training of existing organizational management personnel and improving the management capabilities of the current national fitness volunteer service organizations. The competence of management personnel plays a key role in an organization's efficient operation, necessitating the strengthening of cohesion between management personnel, volunteers, and the organization, along with coordination and cooperation ability between departments. Finally, local sports laws must emphasize the significance of the national fitness volunteer service system, increasing the attention of various levels of departments to the national fitness volunteer service system, and allocating special funds to promote the construction of rules and regulations for the Shanghai national fitness volunteer service system.

Strengthen the Standardization of Training Management Mechanism

The volunteer composition for national fitness exhibits diverse characteristics. In 2009, the General Administration of Sport of China issued a notice on the widespread development of volunteer service activities for national fitness. It divided national fitness volunteers into five categories: excellent athletes; social sports instructors; sports system workers; sports teachers and college students; and community national fitness volunteers.

At present, the training of national fitness volunteers is closely associated with their social identity, such as the systematic and scientific training given to social sports instructors; while other volunteer groups lack the standardized procedures of their training management. To address this concern, research suggests the following methods for the training management mechanism of the standardized national fitness volunteer service system: first, conduct statistical analyses on the current organization and implementation of national fitness activities, along with the existing volunteers' service situation, and establish targeted course training from volunteer types, skill levels, and project content perspectives. Second, when initiating courses on general education training, consider the personalized needs of different types of volunteers, and enhance the diversification of course contents. For instance, volunteers of the coach and instructor types should join specialized training and teaching method training programs, whereas volunteers of the activity assistance type should grasp better the operating process of comprehensive fitness volunteer service activities during the training, with highly pointed course content related to the needs of such activities. Third, institute a reward and punishment mechanism to motivate more personnel to participate in volunteer activities and tackle any existing absence cases from training sessions.

Improve the Standardization of Project Promotion Mechanism

Research suggests that improving the standardization of project promotion mechanisms should begin based on the following aspects reflecting the characteristics of the times: first, strengthening the cooperation between vertical and horizontal governments at all levels and establishing a working group for the national fitness volunteer service to enhance information transmission effectiveness among governments and departments at all levels. Second, broadening the scope of volunteer services for national fitness and intensifying cooperation with volunteer organizations and service departments would enable learning from mature operational experiences of relevant organizations and departments. Furthermore, this approach could expand the sources of national fitness volunteers and promote the inclusion of national fitness volunteer services into the volunteer service sector. Third, enhancing collaboration with schools and enterprises, widening the coverage of national fitness volunteer services, and joint efforts with schools and businesses to develop national fitness volunteers and organize more diverse national fitness volunteer activities would enrich the connotation of national fitness volunteer services while nurturing social participation of the national fitness volunteer service system.

Promote the Standardization of Recognition and Incentive Mechanisms

Based on experience from both domestic and international public sports volunteer service systems, it is clear that the primary way of recognizing and motivating sports volunteers is still through spiritual motivation. While the recognition and incentive mechanism for national fitness volunteers in Shanghai has already begun to take shape, there remains room for improvement. Firstly, in the era of short videos, the form of spiritual motivation for volunteers has diversified. Videos highlighting outstanding individuals and advanced volunteer organizations in the field of national fitness have been posted on relevant communities, national fitness, and volunteer service platforms to promote and encourage further efforts in the field of national fitness volunteer services while also inspiring more people to participate. Furthermore, expanding the scope of recognition and using a wider variety of awards and commendations can enhance the enthusiasm of individuals and organizations for service.

Secondly, greater attention should be paid to the special personal and family circumstances of volunteers. Those with outstanding performance or unique situations may be given necessary material incentives. This represents not only humanistic care for volunteers but also helps ensure the organization's continued operation.

Finally, we can draw from experiences building public sports volunteer service systems in different regions to establish a volunteer level system. Volunteers of different levels can undertake varying service roles, and advancement through long-term service is an effective way to motivate volunteers to actively participate in national fitness volunteer services.

Promoting the Standardization of Evaluation and Inspection Mechanisms

The evaluation and inspection should be clarified based on the targeted subjects, ranging from the organization to the process of national fitness volunteer service. Different objects require corresponding evaluation and inspection mechanisms. Attention should be paid to guiding the indicator system by practice, ensuring scientificity, and using reasonable methods for screening to achieve comprehensive and comprehensive evaluation. Eventually, the national fitness volunteer service system should achieve a high-quality and standardized approach.

In promoting the standardization of the evaluation and inspection mechanism, moderation principles must always be in place. The purpose of evaluation and inspection is not the end goal, but rather a means towards promoting the standardization of the operation mode of Shanghai's national fitness volunteer service. If wrongly regarded as the main objective, evaluation and inspection will dampen the enthusiasm of national fitness volunteer service organizations and volunteers, thus damaging the system.

REFERENCES

Harvey J, Levesque M, & Donnelly P. (2007). Sport volunteerism and social capital. Sociology of Sport Journal, 24(2),206-223.

Findlay-King L, Nichols G, Forbes D, et al. (2008). Watching the pennies and the people - how volunteer-led sport facilities have transformed services for local communities. Managing Sport and Leisure, 23(4-6),277-292.

Kappelides P, & Spoor J. (2009). Managing sport volunteers with a disability: Human resource management implications. Sport Management Review, 22(5), November, 694-707.

Nursey-Bray M, Masud-All-Kamal M, Di Giacomo M, et al. (2022). Building community resilience through youth volunteering: towards a new model. Regional Studies Regional Science, 9(1), December, 242-263.

Rasmussen L G, Skjerning H T, & Burau V. (2023). The why and how of co-production between professionals and volunteers: a qualitative study of community-based healthcare in Denmark. International Journal of Sociology and Social Policy, 43(1/2), February, 197-213.

Sorensen K, Pedersen L S, & Sander J. (2022). Exploring the nexus of health promotion, sport and well-being to improve future synergies and public health capacity through integrated approaches. International Journal of Health Governance, 27(2), May, 143-149.

Wicker P, & Downward P. (2020). The Causal Effect of Voluntary Roles in Sport on Subjective Well-Being in European Countries. Journal of Sport Management, 34(4), July, 303-315.

Wright R K, & Wiersma C. (2020). 'It Pays to Play': The Emergence of Innovative Planning, Occupational Devotion, and Lifestyle Entrepreneurship in Aotearoa/New Zealand. 23-39.

DECISION SCIENCES INSTITUTE

Toward a Conceptual Framework of International Tourism after the Pandemic: Consumption Propensity Based on Expectations of Refreshment and Development

Seung Hoon Jang
Commonwealth University of Pennsylvania
East 2nd Street, Bloomsburg, PA, 17815
Email: sjang@commonwealthu.edu

Christian Grandzol
Commonwealth University of Pennsylvania
East 2nd Street, Bloomsburg, PA, 17815
Email: cgrandzo@commonwealthu.edu

ABSTRACT

This study examines how refreshment and development influence the consumption propensity of international tourists. The study reviews interconnections from the tourism industry, management, and related fields. Theoretical propositions regarding the relationship between refreshment and development with consumption propensity are provided, along with the conceptual model of refreshment and development in the tourism context. The study provides scholars and practitioners with the meaningful insight that international tourists may spend increased amounts due to fulfillment of their expectation of refreshment and the development of tourism attractions. Local governments and businesses should consider these expectations to bolster their local economies after the pandemic.

KEYWORDS: Tourism Industry, Pandemic, Consumption Propensity, International, Travel, Refreshment, Management, Local Business, and Conceptual

INTRODUCTION

There have been pressing concerns about the global economic downturn due to the Coronavirus Disease Pandemic (e.g., Gourinchas, 2020). In this context, scholars and practitioners should examine how to retain and perhaps improve sustainable consumption despite the broader environmental impact of the pandemic. Key to this examination is understanding the tendencies of how and why individuals decide to consume. Propensity to consume is described as “the proportion of total income or of an increase in income that consumers tend to spend on goods and services rather than to save (Encyclopaedia Britannica, 2010).” Propensity to consume implies that individuals can exhibit different levels of consumption due to varied factors. How will consumers continue their spending and contribute to sustainable economic growth despite resource constraints, tremendous uncertainty, travel restrictions, and environmental fluctuations?

Tourism is an industry that helps fulfill the goal of sustainable consumption. Tourism can be described as “leisure travel away from home (Aronsson, 2009, p77).” The tourism industry provides various benefits, such as increased jobs in local regions and the prospects of a better life (Aref & Redzuan, 2009). In other words, tourism contributes to the local economy and community while helping tourists themselves refresh from everyday stress. For example, well-designed local restaurants may increase sales volumes by attracting a portion of tourists visiting the area. It is likely that international tourists expecting benefits such as refreshment and tourism resources to enjoy will be more likely to decide to continue to travel and to spend.

For local economies without sufficient capital and technology, international tourists can provide valuable opportunities for economic maintenance and recovery as capital created elsewhere flows into the local economy. Tourism also helps develop managerial talent, a foundation for future economic development. Finally, tourism enables economic growth while also minimizing negative environmental impact compared to other growth engine – industries such as mining and heavy manufacturing. In this vein, it behooves local leaders to understand how to attract more international tourists.

This article focuses on international travelers' expectations of refreshment and development as key factors of tourism consumption propensity encouraging local businesses and jobs after the pandemic. To accomplish this objective, the literature of tourism management and related disciplines is reviewed. Theoretical propositions are provided based on the review and discussion. In addition, the conceptual model of refreshment and development is provided to help researchers interested in this topic. Limitations and contributions of this study are discussed at the end of this study. Future scholars and practitioners are expected to gain implications on how to attract international tourists by focusing on their expectations of refreshment and development related to tourism attractions.

LITERATURE REVIEW AND CONCEPTUAL DEVELOPMENT

International tourism and propensity to consume

The tourism industry is considered one of the main sources of economic growth and prosperity (e.g., Costa et al., 2020). As described in Table 1, studies have investigated various environmental and other factors influencing tourism businesses (see Table 1). There has also been growing interest on how to manage an environmentally-sustainable tourism industry (Pigram, 1996), as interest has increased in sustainability as well as economic effects related to tourism activities. Therefore, it is important for scholars and practitioners to identify ways of achieving these oftentimes conflicting objectives. Scholars have focused on factors explaining the degree of consumption in the tourism industry (e.g., Wu et al., 2016). Richards (2002) wrote that "tourists are motivated to visit a destination by information received from a generating marker that matches their needs and wants (p 1050)." Lepp and Gibson (2003) examined the role of novelty and familiarity of international tourists in explaining their perceived risk of international tourism. As a seminal study on bounded rationality (March, 1978) implies, individual decision making can be influenced by various factors other than rational consideration. Therefore, it is important to investigate factors other than economic conditions to understand how and why consumers decide to spend for international travel.

Table 1: Relevant studies on tourism

Articles	Contents
Aronsson, L. (2009)	Tourism can be described as "leisure travel away from home (Aronsson, 2009, p77)."
Choi & Sirakaya (2005)	A SUS-TAS scale on sustainable community tourism was designed and empirically tested.
Gössling et al. (2020)	This study examines how COVID-19 pandemic influences tourism as well as economy and society.
Pigram (1996)	Best practices and auditing of environment friendly tourism were discussed.
Richards (2002)	Examined the relations among factors like tourists' markers, characteristics, motivation, and tour decision conceptually and empirically.

According to Vergauwe and Langerock (2017), refreshment is “an active, attention-based maintenance process in working memory (p. 23)”. Although less focus has been on the influence expectation of refreshment on tourism businesses in our understanding, refreshment is also one of the main dimensions that drives consumers in the tourism setting (Kim et al., 2012). Tourists may refresh their lives by visiting foreign tourism attractions, experiencing local cultures, and simply being away from their routine schedules. As the focus on novel consumer satisfaction (Bianchi, 1998) implies, individuals may enjoy newness and recharge vitality during their visit to international places and communities. Individuals also tend to prefer different, but friendly, attractions and cultures to refresh with less negative feelings toward unexpected features. For example, tourists may not consider unexpected, overly-priced tickets as refreshing; rather this could be a source of stress. Thus, it is important that tourists expect that they will refresh themselves through their tourism. It is also probable that tourists’ positive expectations for tourism attractions encourage individuals to visit and enjoy certain destinations. For example, people who read local history books may be more willing and eager to visit local ancient relics as a source of refreshment. When refreshment expectations are not met, individuals will be less likely to consider international travel.

International tourism businesses can recognize the level and/or sustainability of expected refreshment of potential travelers by utilizing refreshment measures. If individuals of certain countries or regions expect a high level of initial or sustained refreshment from travel packages, tourism businesses need to promote their travel products actively in the areas. For example, if children watch a movie located in New York City, they begin to have recognition of famous local attractions such as the Empire State Building and the Statue of Liberty. In the future, they may consider tourism packages including the local attractions based on such memories and images. Thus, countries might expect increased international tourism consumption propensity when travelers are assured of refreshment.

Tourists can also derive refreshment from other sources during international travel such as duty-free products when considering foreign attractions. For instance, individuals can purchase nice sweaters at special prices in the airport duty free shop. These amenities generate additional capital for the local area while also likely increasing consumption propensity. Other factors such as differing climates can also encourage international tourists to expect refreshment. While tourists from warmer countries may travel to a foreign country with the expectations of refreshing experiences such as skiing and snow tubing, people in colder areas may wish to be refreshed by warm conditions.

Therefore, the following proposition is suggested:

Proposition 1 International tourists are more likely to spend more for and during their travel when they expect a higher degree of refreshment.

In addition to refreshment, international tourists may consider the degree of expected development of tourism products and services when choosing their destinations, especially because overseas travel requires more time, documentation, budget, and perhaps risk than domestic travel. Thus, international tourists are often more sensitive to the quality and contents of tourism products and services. Tourists may also spend more for tourism packages if they believe tourism development has been pursued in harmony with the surrounding environment. For instance, individuals might consider the types of developed facilities such as observatory elevators. Those expecting well-developed tourism products and services are more likely to plan to spend for the additional opportunities. For example, certain tourists may reserve a private hot spring and/or traditional cuisine options despite the costs associated with them. Thirsty tourists engaged in trail hiking are more likely to purchase a beverage at a nearby cafe.

Local governments and businesses can promote such tourism attractions (e.g., Marzuki, 2010). Potential tourists might watch commercials on a new observatory deck near a famous waterfall along with small business districts that provide additional shopping opportunities and decide to travel to the destination. Local governments and business groups often take the lead in developing tourism attractions. First, existing tourism resources are physiologically developed. Local governments may, for example, build an airport, a train station, or highways to improve accessibility and convenience of tourism attractions. As a result, tourists might be more likely to visit the locale and in so doing generate additional local tourism revenues. For example, the construction of a new airport that enables tourists to visit the surrounding areas in 2 hours rather than 6 hours via conventional routes could increase a traveler's likelihood to visit. Second, development can be implemented from an intangible perspective. For example, a new promotion team may focus on a social-friendly safari tour in which a portion of profits are used to support local poverty reduction or protect animal habitats.

Tour companies can also develop different types of intangible resources associated with current situations. For examples, there have been discussions on how to deal with Covid-19 pandemic issues (e.g., Gourinchas, 2020). If tourism businesses can persuade potential tourists that they will enjoy scenery and attractions with social distancing and proper precautions at the location, this expectation may encourage tourists to travel. In addition to conventional mediums such as booklets and agency advertisement, travel companies can utilize other channels of introducing new tourism developments such as social media networks. People may see an area's update regarding newly developed tangible or intangible resources, access other opinions of related development, or witness others in their network who have visited the location. Virtual platforms also enable individuals to evaluate local scenery through videos and virtual tours.

It is also important to provide varied options of tourism products for international tourists given people's unique interests and needs. For example, individuals are more likely to understand the benefits and characteristics of local traditional cuisine if descriptions and instructions are communicated in their native language. Local governments or stakeholders may provide translated versions of poems or novels related to their historic attractions. The focus of development of tourism products is how to drive individual traveler's intention to consume both products and services. Overall, it is probable that local businesses can attract global tourists and the related consumption by developing tangible and intangible tourism resources and then properly introducing them to potential tourists.

Therefore, the second proposition is suggested as follows:

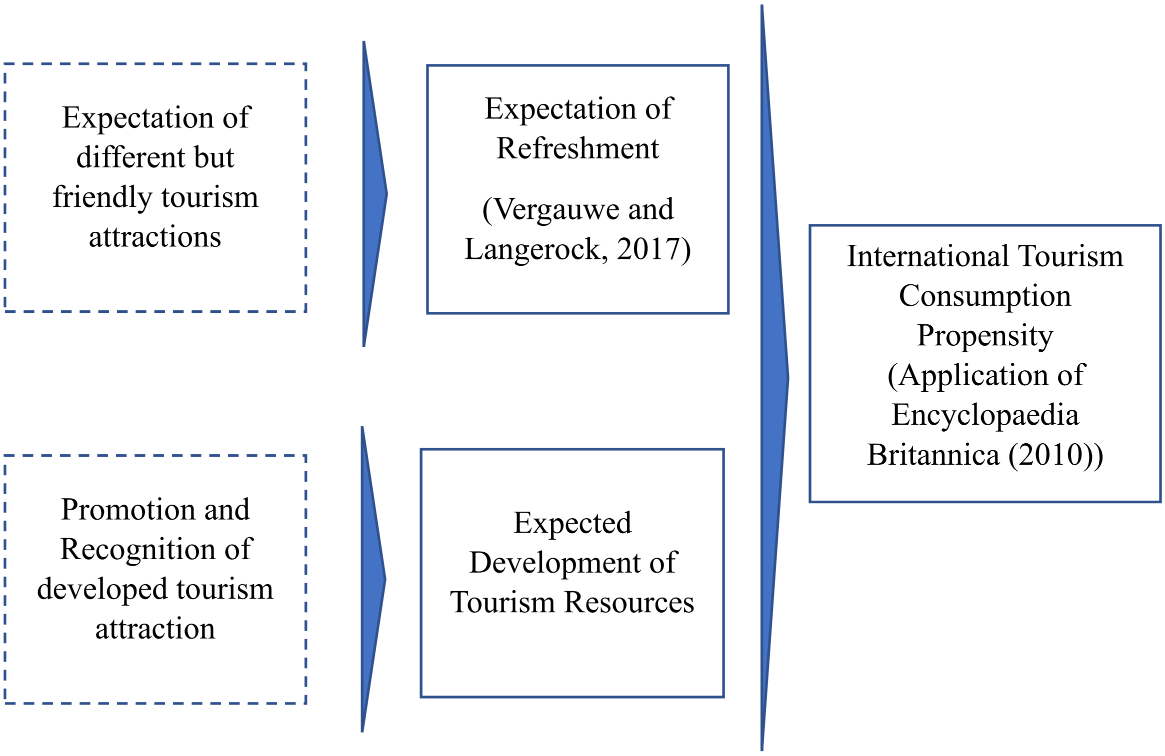
Proposition 2 International tourists are more likely to spend for their travel when they expect highly-developed tourism products and services that align with their interests.

Based on the discussion above, the authors provide the function and the conceptual framework (Figure 1) of tourism consumption propensity as follows:

International Tourism Consumption Propensity =

F (Expectation of Refreshment, Expected Development of Tourism Resources, etc.)

Figure 1: Conceptual framework of International Tourism Consumption Propensity



*Based on the discussion in the main body of this article

These function and framework posit that tourists are more likely to consume if they expect refreshment from international travelling and development of the tourism location and attractions. It also implies that tourists traveling to new attractions in foreign countries will spend more than those visiting familiar domestic attractions due to increased expectations of refreshment and novel experiences brought by development. Individual tourists can be attracted to easy-to-understand opportunities to experience foreign culture, new attractions, and local food.

LIMITATIONS AND DISCUSSIONS

Scholars and practitioners can obtain valuable implications from this study. First, we recommend that leaders in the tourism industry focus on the effects of refreshment on the consumption of international travelers. Local governments developing tourism attractions should also consider means to attract foreign tourists to maximize economic effects on their local economies. As an example, subway signs in South Korea and Japan with neighboring countries' languages like Korean, Japanese, and Chinese as well as English would help make hosting tourists from overseas more likely.

Secondly, scholars in this research stream should pay attention to the role of perceived development of attractions and related products and services in explaining consumption propensity of tourists. Even if individuals expect a similar level of refreshment from the same location, actual consumption might vary widely due to differently developed products and services. A tourist visiting two similar waterfalls, for example, might spend more at the one with well-designed tourism packages or nice restaurants with fantastic scenery.

The suggested models of refreshment and development enable scholars and practitioners to examine these issues theoretically. Future scholars are recommended to examine the influence of refreshment and development related to the tourism industry systematically. It can broaden the understanding of how to attract international tourists and boost local economies. Even neighboring countries might attract tourists from each other and realize increased consumption based on expected refreshment and development, resulting in mutually beneficial relations. The function and conceptual framework may also help scholars interested in other industries examine and understand the role that refreshment and development have on them.

CONCLUSION

The tourism industry has been significantly influenced and negatively affected by the global pandemic (Gössling et al., 2020). This phenomenon vividly reminded many people of the importance of the tourism industry for employment, small businesses, and local government revenues. As discussed above, tourism is one way of increasing jobs and boosting local economies. As tourists visit an area, it results in spending for transportation, lodging, meals, shopping, and other services that support a variety of workers. This study broadens the understanding of this important industry by providing theoretical propositions based on refreshment and development.

Given that international tourists tend to spend more money, facilitating international tourism and enhancing tourism consumption propensity result in mutually beneficial economic gains and cooperation among countries. It implies that each country can expect increased tourism consumption when a focus is on sharing tourists with other countries. The global pandemic also reminded people that we depend on each other and must find win-win scenarios among foreign countries for the best benefit of all. This study contributes to the discussion on this topic via a theoretical review and discussion and introduction of the conceptual model of refreshment and development.

REFERENCES

Aref, F., & Redzuan, M. (2009). Community leaders' perceptions toward tourism impacts and level of community capacity building in tourism development. *Journal of Sustainable Development, 2*(3), 208-213.

Aronsson, L. (2009). Sustainable tourism systems: The example of sustainable rural tourism in Sweden. *Journal of Sustainable Tourism, 2*(1-2), 77-92.

Bianchi, M. (1998). Consuming novelty: Strategies for producing novelty in consumption. *Journal of Medieval and Early Modern Studies, 28*(1), 3-18.

Choi, H-. S. C., & Sirakaya, E. (2005). Measuring residents' attitude toward sustainable tourism: Development of sustainable tourism attitude scale. *Journal of Travel Research, 43*(4), 380-394.

Costa, B. J. F, Rodrigues, S. C. S. F., & Pacheco P. M. (2020). Circular economy and the tourism industry. *Journal of Global Business and Technology, 16*(1), 45-61.

Encyclopaedia Britannica. (2010). *Propensity to consume,* Encyclopaedia Britannica. Retrieved from https://www.britannica.com/topic/propensity-to-consume , May 23, 2020.

Gourinchas, P. O. (2020). *Flattening the pandemic and recession curves*. in Baldwin, R., & Mauro, B. W. (Ed.): *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever,* 31-39. London: CEPR press.

Gössling, S., Scott, D., & Hall, C. (2020). Pandemics, tourism and global change: A rapid assessment of Covid-19. *Journal of Sustainable Tourism, 29*(1), 1-20.

Kim, J-. H., Ritchie, J. R. B., & McCormick, B. (2012). Development of a scale to measure memorable tourism experiences. *Journal of Travel Research, 51*(1), 12-25.

Lepp, A., & Gibson, H. (2003). Tourist roles, perceived risk and international tourism. *Annals of Tourism Research, 30*(3), 606-624.

March, J. G. (1978). Bounded rationality, ambiguity, and the engineering of choice. *Bell Journal of Economics, 9*(2), 587-608.

Marzuki, A. (2010). Tourism development in Malaysia.: A review on federal government policies. *Theoretical and Empirical Researches in Urban Management, 5*(8), 85-97.

Pigram, J. J. (1996). Best practice environmental management and the tourism industry. *Progress in Tourism and Hospitality Research, 2*(3-4), 261-271.

Richards, G. W. (2002). Tourism attraction systems: Exploring cultural behavior. *Annals of Tourism Research, 29*(4), 1048-1064.

Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory Language, 96*, October, 23-35.

Wu, K., Raab, C., Chang, W., & Krishen, A. (2016). Understanding Chinese tourists' food consumption in the United States. *Journal of Business Research, 69*(10), 4706-4713.

DECISION SCIENCES INSTITUTE

Toward a Normative Framework for Systemic Risk Modeling

Faraz Dadgostari
Montana State University
faraz.dadgostari@montana.edu

Peter Beling
Virginia Tech
beling@vt.edu

ABSTRACT

In this paper, we present a mathematical formulation of financial systems that models an interbank network within the context of the real economy. Our aim is to measure and optimize systemic risk through a dynamic parameter setting mechanism. We construct a mathematical model as the foundation for designing a regulatory framework. Furthermore, we extend this model to a decision-making/optimization model, where the systemic parameters are directly set as decision variables for systemic risk management. The proposed model leverages the systemic structure of contractual dependencies among financial institutions, taking into account the banks' interests in maximizing their shares of the financial market.

KEYWORDS: Network modeling, Systemic risk, Dynamic parameter setting

DECISION SCIENCES INSTITUTE

Understanding Students' Learning Style Transformation through Audit Simulation: A MENA Region University Case

Nader Elsayed
Qatar University
Email: nelsayed@qu.edu.qa

Mostafa Kamal Hassan
Qatar University
Email: mostafa.hassan@qu.edu.qa

ABSTRACT

Drawing on Experiential Learning Theory, this study reveals how students' learning styles transformed through Audit Simulation (AS) experience in a MENA university for sustaining students' lifelong learning. In our case study, 48 participants completed independently a questionnaire of six open-ended questions; this information is paired with follow-up feedback and instructor's observations. The study finds that the AS assignment transformed students' learning by enabling them to experience different learning styles. They visualised an authentic AS experience in critically analysing and practically evaluating AS documents. The study has several implications for the accounting education literature, professional accounting bodies, and educators.

KEYWORDS: Audit Simulation, Experiential Learning Theory, Accounting Education, Case Study, MENA.

DECISION SCIENCES INSTITUTE

Transportation Problem with Multiple Trucks and Constraints on Capacity and Budget

Hadi Farhangi
Savannah State University
Email: farhangih@savannahstate.edu

Abel Hafize Ouedraogo
Savannah State University
Email: aouedra2@student.savannahstate.edu

ABSTRACT

This work extends the transportation problem to account for multiple trucks. It considers two scenarios for this extension. The first scenario assumes a manager needs to purchase trucks to deliver the goods in the settings of the transportation problem. The manager needs to consider the limitation on the budget and the capacity of trucks. The second scenario assumes the manager already has a fleet of trucks available and has to decide how to transport using the fleet of trucks. This study formulates both scenarios and analyzes how the parameters of the new problems will affect the truck selection and truck loads.

KEYWORDS: Transportation Problem, Trucks, Parameter Analysis

INTRODUCTION

Transportation is an essential component of any economy, enabling goods and services to move efficiently from suppliers to customers. This problem, in its original form, decides the quantity to transport from multiple origins to multiple destinations while minimizing the cost of transportation (Bazaraa et al., 2011).

This problem becomes particularly complex when being used in practice. For instance, the work of Díaz-Parra et al. (2014) provides a survey on air, maritime, railroad, modal, space, and school bus transportation. Another example is using multiple modes of transportation in this problem (Maity et al. 2020; Flórez et al., 2011). Work of Balaji & Balaji (2018) considers multiple periods of transportation in this problem. For more variants and applications, one may read (Díaz-Parra et al., 2014; Bookbinder & Sethi, 1980).

This paper, however, considers a scenario of transportation problem when a manager has to decide how many trucks to purchase to deliver goods. The problem, then, faces limited budget and truck capacity. The budget limitation in the transportation problem is studied as a bottleneck transportation problem (Charnsethikul & Svetasreni, 2007). However, to the best of the authors' knowledge, there is no work that considers multiple trucks, budget, and truck capacity simultaneously. Another similar scenario in the transportation problem arises when a manager already has a fleet of trucks and needs to assign these trucks to carry out the transportation of goods. This problem also is not addressed in the literature and it is the second contribution of this paper.

These two scenarios are formulated as extensions to the transportation problem. For both of these problems the parameters such as number of suppliers, number of destinations, the number of trucks, capacity of trucks, total demand, and budget are studied to see their effects on the number of selected trucks and their loads. The formulations of these two problems are presented in the next section. The analysis on the range of parameters is presented afterward. Paper then presents a discusses on the findings of the numerical analysis and the final section concludes the paper.

PROBLEM FORMULATION

This section formulates the two scenarios discussed in the previous section. The first scenario deals with the transportation problem when managers have to purchase trucks when having limited budgets and limited truck capacity. To formulate the problem, three sets are defined. Set I contains all the suppliers and it is indexed by $i \in I$. Set J contains all the suppliers and it is indexed by $j \in J$. Set T contains all the trucks and it is indexed by $t \in T$. It is assumed that trucks have the same capacity.

The problem includes six different parameters. Parameter s_i is the supply of supplier $i \in I$. Parameter d_j is the demand at destination $j \in J$. Parameter c_{ij} is the cost of transportation between supplier $i \in I$ and destination $j \in J$. Parameter c^t is the cost of purchasing truck $t \in T$. Parameter B is the available budget and parameter cap shows the capacity of trucks.

There are two types of variables in the formulation. Particularly, variable x_{ij}^t shows the amount of transportation from supplier $i \in I$ to destination $j \in J$ by truck $t \in T$. In addition, binary variable $z^t = 1$, if truck $t \in T$ is used. Otherwise, variable z^t is set to zero. Given these sets, parameters, and variables, the problem is formulated as follows:

$$P_1: \quad \min \quad \sum_{t \in T} c^t z^t \tag{1}$$

$$\text{s.t.} \quad \sum_{t \in T} \sum_{j \in J} x_{ij}^t \le s_i \qquad \forall i \in I \tag{2}$$

$$\sum_{t \in T} \sum_{i \in I} x_{ij}^t \ge d_j \qquad \forall j \in J \tag{3}$$

$$M z^t \ge \sum_{i \in I} \sum_{j \in J} x_{ij}^t \qquad \forall t \in T \tag{4}$$

$$x_{ij}^t \le cap \qquad \forall t \in T, \forall i \in I, \forall j \in J \tag{5}$$

$$\sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij}^t + \sum_{t \in T} c^t z^t \le B \tag{6}$$

$$x_{ij}^t \ge 0 \qquad \forall t \in T, \forall i \in I, \forall j \in J \tag{7}$$

$$z^t \in \{0,1\} \qquad \forall t \in T \tag{8}$$

Equation (1) is the mathematical format for the objective function that minimizes purchasing cost of trucks. Constraints (2) are the mathematical formulation of a constraint that specifies that the total amount of transportation from a source i to all destinations j cannot exceed the supply s_i available at supplier i. Constraints (3) specify that the total amount of transportation from all sources i to a destination j must be at least equal to the demand d_j at destination j. Constraints (4) force the selection of truck t if there is a transportation by this truck. Here, M is a large positive number. Constraint (5) enforce the same capacity for all trucks. Constraint (6) is the bottleneck constraint. It enforced the total cost of transportation and truck purchasing to be less than the available budget. Constraints (7) and (8) define decision variables.

The second scenario is similar to problem P_1. However, the manager have a fleet of trucks available with limited capacity, and they need to minimize the total operation cost (the cost of truck operation and the cost of transportation). The same notation is used to formulate this scenario, with one difference that the parameter c^t represents the operational cost of truck t instead of its purchasing cost.

$$P_2: \quad min \quad \sum_{t \in T} \sum_{i \in I} \sum_{j \in J} c_{ij} x_{ij}^t + \sum_{t \in T} c^t z^t \tag{9}$$

$$s.t. \quad \text{Constraints (2), (3), (4), (5), (7), (8)}$$

Equation (9) minimizes the total cost of transportation and truck operation. The rest of constraints are borrowed from problem P_1. Note that constraint (6) is not included in problem P_2.

NUMERICAL ANALYSIS

Problems P_1 and P_2 are solved and analyzed in this section. These problems are implemented in Python using the MIP package. The CBC solver is utilized by the MIP package to return the optimal solution. All codes are run on a computer with five 3.4 GHz core CPUs and 8 GB RAM.

To perform the analysis, the values of parameters, such as demand, capacities, number of suppliers, number of destinations, and available budget are varied and the results are recorded. Originally, parameters I, J, T are generated as integers between 2 and 10. Parameter s_i is between 1000 units and 2000. Parameter d_j is between 100 units and 200. The transportation cost (c_{ij}) is between \$5 and \$20. The truck purchasing cost (c^t) is between \$30,000 and \$100,000. Truck capacity (cap) varies between 50 units and 100 units. The “Demand” in tables shows the total demand in the solved instance and “CPU” shows the completion time to solve the instances in seconds.

Analyzing problem P_1

Given the range of parameters defined in the beginning of this section, eight problem instances are solved and reported in Table 1. Note that the budget is calculated as half of the sum of all transportation costs and truck purchasing costs. In this table, the number of variables, number of constraints, capacity, budget, total demand, number of trucks, load of trucks, and computational time (CPU) are reported. All these instances are solved in less than a second. Note that larger problem sizes are intentionally avoided. In all these instances, either one truck or two trucks are purchased. When two trucks are purchased, their load do not differ drastically, instead, they are balanced. Note that the objective of problem P_1 is to reduce the cost of truck purchase. The model, then, finds the minimum number of trucks that have lowest costs and assigns transportation to these trucks. Because of this, the main difference between transportation problem and P_1 is the number of trucks and loads of these trucks. In what follows, most of parameters of problem will vary to study truck the effects of them on the number of purchased trucks and the truck loads in problem P_1. These parameters are capacity, budget, demand, number of suppliers, number of destinations, and number of available trucks.

Table 1. Analyzing problem P_1

I	J	T	# var	# constraints	Capacity	Budget	Demand	# of trucks	Trucks Load	CPU
2	6	6	78	87	50	211606	949	2	[520,429]	0.012
2	8	6	102	113	50	201342	1192	2	[532,660]	0.017
2	10	6	126	139	50	185746	1596	2	[763,833]	0.045
2	6	8	104	113	50	243126	767	2	[365,402]	0.017
2	6	10	130	139	50	380918	943	2	[474,469]	0.016
4	6	6	150	161	50	183112	966	1	966	0.014
4	8	6	198	211	50	183697	1258	1	1258	0.03
4	10	6	246	261	50	161278	1368	1	1368	0.02

<u>Effects of truck capacity on truck selection and loads</u>

Here, the capacity of trucks is varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J\in\{6,8,10\}$, $T\in\{6,8\}$, the budget is fixed as the beginning of the section dictates its calculation, and demands are stated as necessary. Tables 2, 3, 4, and 5 summarize the results of varying truck capacity between 50 and 100 units. Table 2 shows the result of analysis, when $I=2$, $J=6$, $T=6$, $D=900$, and $B=\$181{,}149$. This table shows that once the capacity started to increase, the number of trucks either remains the same or decreases. This is a similar trend in Tables 3, 4, and 5. In all these tables, as the truck capacity increases, the number of selected trucks either remains the same or decreases. This is expected because an increase in capacity will allow a truck to have higher loads. Another observation is that the loads of trucks are quite close to each other.

Table 2. I=2, J=6, T=6, D=900, B=$181,149

Capacity	# of trucks	Trucks Load	CPU
50	2	[484,416]	0.02
60	2	[433,467]	0.02
70	2	[413,487]	0.02
80	2	[412,488]	0.02
90	1	900	0.01
100	1	900	0.01

Table 3. I=2, J=8, T=6, D=1330, B=$201,327

Capacity	# of trucks	Trucks Load	CPU
50	2	[589,741]	0.01
60	2	[547,783]	0.02
70	2	[677,653]	0.02
80	2	[655,675]	0.04
90	2	[605,725]	0.03
100	1	1330	0.01

Table 4. I=2, J=10, T=6, D=1705, B=$185,249

Capacity	# of trucks	Trucks Load	CPU
50	2	[823,882]	0.03
60	2	[823,882]	0.02
70	2	[744,961]	0.02

Table 5. I=2, J=6, T=8, D=803, B=$277,730

Capacity	# of trucks	Trucks Load	CPU
50	2	[403,400]	0.02
60	2	[386,417]	0.01
70	2	[386,417]	0.01

80	2	[753,952]	0.02
90	2	[833,942]	0.02
100	1	1705	0.01

80	2	[386,417]	0.01
90	1	803	0.01
100	1	803	0.02

Effects of budget on truck selection and loads

Here, the available budget is varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J \in \{6,8,10\}$, $T \in \{6,8\}$, the capacity is fixed at 50 units, and demands are reported when we have a feasible problem. Tables 6, 7, 8, and 9 summarize the results of varying the budget. The budget is changing from the quarter of the total sum of truck costs and transportation to the half of it. The increase is at 10% level. Table 6 shows that once the budget increases, the number of selected trucks and their load will not change. This is the same in Tables 7, 8, and 9. Based on these tables, as the budget increases, the number of selected trucks remains the same. This is expected because an increase in budget will not enforce a more limitations on the problem. Another observation is that the loads of trucks are quite close to each other.

Table 6. I=2, J=6, T=6, cap=50, D=930

Budget	# of trucks	Trucks Load	CPU
88731	2	[459,471]	0.01
106477	2	[459,471]	0.01
124223	2	[459,471]	0.01
141970	2	[459,471]	0.01
159716	2	[459,471]	0.01

Table 7. I=2, J=8, T=6, cap=50, D=1320

Budget	# of trucks	Trucks Load	CPU
109804	2	[598,722]	0.01
131765	2	[598,722]	0.01
153726	2	[598,722]	0.01
175687	2	[598,722]	0.02
197648	2	[598,722]	0.01

Table 8. I=2, J=10, T=6, cap=50, D=1445

Budget	# of trucks	Trucks Load	CPU
99309.2	2	[818,621]	0.02
119171	2	[720,725]	0.02
139033	2	[720,725]	0.02
158895	2	[720,725]	0.02
178757	2	[720,725]	0.02

Table 9. I=2, J=6, T=8, cap=50, D=967

Budget	# of trucks	Trucks Load	CPU
125462	2	[502,465]	0.02
150555	2	[502,465]	0.02
175648	2	[502,465]	0.02
200740	2	[502,465]	0.01
225832	2	[502,465]	0.01

Effects of demand on truck selection and loads

Here, the total demand is varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J \in \{6,8,10\}$, $T \in \{6,8\}$, the capacity is fixed at 50 units, and the budget is set to be as twice the value reported n the beginning of this section to support the analysis of demand increase. Tables 10,11,12, and 13 summarize the results of varying demand. The demand is changing from what it is originally calculated, in the beginning of this section, by 10 units for every destination at a time. The analysis stops when the total demand reaches half of the total supply or enough observations are made. Based on these tables, as the demand increases, the number of selected trucks either remains the same or increases. This is expected because an increase in demand requires more trucks to fulfill the demand. Another observation is that the loads of trucks are quite close to each other.

Table 10. I=2, J=6, T=6, cap=50, B=$360,387

Demand	# of trucks	Trucks Load	CPU
816	2	[421,395]	0.01
876	2	[441,435]	0.02
936	2	[458,478]	0.02
996	2	[600,600]	0.01
1056	3	[323,333,400]	0.02
1116	3	[338,363,415]	0.03
1176	3	[348,393,435]	0.03

Table 11. I=2, J=8, T=6, cap=50, B=$408,701

Demand	# of trucks	Trucks Load	CPU
1177	2	[519,658]	0.02
1257	2	[569,688]	0.02
1337	3	[351,502,484]	0.05
1417	3	[401,512,504]	0.04
1497	3	[439,522,536]	0.05

Table 12. I=2, J=10, T=6, cap=50, B=$415,212

Demand	# of trucks	Trucks Load	CPU
1311	2	[702,609]	0.03
1411	2	[740,671]	0.02
1511	2	[770,741]	0.02
1611	3	[532,583,496]	0.05

Table 13. I=2, J=6, T=8, cap=50, B=$443,244

Demand	# of trucks	Trucks Load	CPU
904	2	[485,419]	0.02
964	2	[530,434]	0.02
1024	3	[275,349,400]	0.03
1084	3	[296,379,409]	0.03

Effects of suppliers, destinations, and available trucks on truck selection and loads

Here, the number of suppliers, destinations, and available trucks is being varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J\in\{6,8,10\}$, $T\in\{6,8\}$, demands are calculated as the beginning of this section dictates, the capacity is fixed at 50 units, and the budget is set to be $200,000. Tables 14,15,16,17 summarize the results of the study.

Based on these tables, as the number of suppliers increases, the number of selected trucks decreases. This is because new transportation is added to the model and because the budget is fixed, the model should compensate for the additional transportation by dropping a truck. This, however, is not observed when the number of destinations is increased. In addition, when the number of available trucks increases from 2 to 8 (each table corresponds to a certain number of trucks), there is no change in the number of selected trucks. Another observation is that the loads of trucks are quite close to each other.

Table 14. T=2

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	338	2	[200,138]	0
2	4	629	2	[378,251]	0
2	6	760	2	[376,384]	0
2	8	1243	2	[591,652]	0
4	2	320	1	320	0

Table 15. T=4

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	389	2	[200,189]	0.01
2	4	567	2	[321,246]	0.01
2	6	937	2	[445,492]	0.01
2	8	1177	2	[537,640]	0.02
4	2	259	1	259	0.01

4	4	564	1	564	0
4	6	887	1	887	0
4	8	1269	1	1269	0.01

4	4	665	1	665	0.01
4	6	922	1	922	0.01
4	8	1296	1	1296	0.01

Table 16. T=6

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	309	2	[186,123]	0.01
2	4	566	2	[336,230]	0.01
2	6	1074	2	[540,534]	0.01
2	8	1179	2	[485,694]	0.02
4	2	259	1	259	0.01
4	4	635	1	635	0.01
4	6	889	1	889	0.01
4	8	1104	1	1104	0.02

Table 17. T=8

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	360	2	[200,160]	0.01
2	4	622	2	[351,271]	0.01
2	6	891	2	[447,444]	0.02
2	8	1257	2	[528,729]	0.02
4	2	330	1	330	0.01
4	4	665	1	665	0.01
4	6	957	1	957	0.02
4	8	1213	1	1213	0.03

Analyzing problem P_2

Given the range of parameters defined in the beginning of this section, eight problem instances are solved and reported in Table 18. Note that the budget is calculated as half of the sum of all transportation costs and truck purchasing costs. In this table, the number of variables, number of constraints, capacity, budget, total demand, number of trucks, load of trucks, and computational time (CPU) are reported similar to Table 1. All these instances are solved in less than a second. In all these instances, two trucks are purchased and their loads do not differ drastically, instead, they are balanced. Note that the objective of problem P_2 is to reduce the total cost of transportation and the cost of truck purchase. In what follows, most of parameters of the problem will vary to study the effects of these parameters on the number of purchased trucks and the truck loads. These parameters are capacity, demand, number of suppliers, number of destinations, and number of available trucks.

Table 18. Analyzing problem P_2

I	J	T	# var	# constraints	Capacity	Demand	# of trucks	Trucks Load	CPU
2	6	6	78	87	50	982	2	[437,545]	0.034
2	8	6	102	113	50	1245	2	[660,585]	0.036
2	10	6	126	139	50	1501	2	[788,713]	0.064
2	6	8	104	113	50	816	2	[417,399]	0.039
2	6	10	130	139	50	789	2	[391,398]	0.535
4	6	6	150	161	50	923	1	923	0.113
4	8	6	198	211	50	1084	1	1084	0.028
4	10	6	246	261	50	1558	1	1558	0.085

<u>Effects of truck capacity on truck selection and loads</u>

Here, the capacity of trucks is varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J \in \{6,8,10\}$, $T \in \{6,8\}$, and demands are reported. Tables 19,20,21,22 summarize the results of varying truck capacity between 50 and 100 units. As the truck capacity increases, the number of selected trucks either remains the same or decreases. This is expected because an increase in capacity will allow a truck to have higher loads. Another observation is that the loads of trucks are quite close to each other.

Table 19. I=2, J=6, T=6, D=976

Capacity	# of trucks	Trucks Load	CPU
50	2	[480,496]	0.02
60	2	[480,496]	0.03
70	2	[480,496]	0.02
80	2	[480,496]	0.02
90	2	[470,506]	0.02
100	1	976	0.01

Table 20. I=2, J=8, T=6, D=1238

Capacity	# of trucks	Trucks Load	CPU
50	2	[607,631]	0.06
60	2	[487,751]	0.07
70	2	[500,738]	0.09
80	2	[523,715]	0.08
90	1	1238	0.03
100	1	1238	0.02

Table 21. I=2, J=10, T=6, D=1630

Capacity	# of trucks	Trucks Load	CPU
50	2	[813,817]	0.05
60	2	[788,842]	0.07
70	2	[721,909]	0.08
80	2	[781,949]	0.08
90	2	[827,803]	0.07
100	1	1630	0.02

Table 22. I=2, J=6, T=8, D=834

Capacity	# of trucks	Trucks Load	CPU
50	2	[430,404]	0.05
60	2	[490,344]	0.05
70	2	[452,382]	0.06
80	2	[433,401]	0.05
90	1	834	0.02
100	1	834	0.02

<u>Effects of demand on truck selection and loads</u>

Here, the total demand is varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J \in \{6,8,10\}$, $T \in \{6,8\}$, and the capacity is fixed at 50 units. Tables 23,24,25,26 summarize the results of varying demand. The demand is changing from what it is originally calculated by 10 units for every destination at a time. The analysis stops when the total demand reaches half of the total supply or enough observations are made. Based on these tables, as the demand increases, the number of selected trucks either remain the same or increases. This is expected, because an increase in demand requires more trucks to fulfill the additional demand. Another observation is that the loads of trucks are quite close to each other.

Table 23. I=2, J=6, T=6, cap=50

Demand	# of trucks	Trucks Load	CPU
874	2	[379,495]	0.04
934	2	[519,415]	0.04
994	2	[509,485]	0.03

Table 24. I=2, J=8, T=6, cap=50

Demand	# of trucks	Trucks Load	CPU
1208	2	[637,571]	0.04
1288	3	[390,448,450]	0.04
1368	3	[440,478,450]	0.06

1054	2	[549,505]	0.03
1114	3	[457,350,307]	0.08
1174	3	[448,400,326]	0.08

1448	3	[497,451,500]	0.04
1528	3	[548,471,509]	0.04
1608	3	[487,591,530]	0.04

Table 25. I=2, J=10, T=6, cap=50

Demand	# of trucks	Trucks Load	CPU
1259	2	[577,682]	0.08
1359	2	[657,702]	0.05
1459	2	[687,772]	0.06
1559	2	[767,792]	0.05
1659	3	[533,526,600]	0.09
1759	3	[563,556,640]	0.06

Table 26. I=2, J=6, T=8, cap=50

Demand	# of trucks	Trucks Load	CPU
952	2	[473,479]	0.04
1012	3	[254,415,343]	0.06
1072	3	[282,377,413]	0.05
1132	3	[312,395,425]	0.05
1192	3	[342,415,435]	0.04
1252	3	[379,378,495]	0.05

Effects of suppliers, destinations, and available trucks on truck selection and loads

Here, the number of suppliers, destinations, and available trucks is being varied to study the effects it may have on the number of selected trucks and their loads. For this particular numerical study, $I=2$, $J \in \{6,8,10\}$, $T \in \{6,8\}$, demands are calculated as the beginning of this section dictates, and the capacity is fixed at 50 units. Tables 27,28,29,30 summarize the results of this analysis.

Based on these tables, as the number of suppliers increases, the number of selected trucks decreases. This is similar to problem P_1. It may be because new transportation is added to the model and because the budget is fixed, the model should compensate for the additional transportation by dropping a truck. This, however, is not observed when the number of destinations is increased. In addition, when the number of available trucks increases from 2 to 8 (each table corresponds to a certain number of trucks), there is no change in the number of selected trucks. Another observation is that the loads of trucks are quite close to each other.

Table 27. T=2

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	249	2	[149,100]	0.01
2	4	524	2	[274,250]	0.01
2	6	913	2	[525,388]	0.01
2	8	1169	2	[586,583]	0.01
4	2	328	1	328	0.01
4	4	648	1	648	0.01
4	6	894	1	894	0.02
4	8	1270	1	1270	0.01

Table 28. T=4

I	J	Demand	# of Trucks	Trucks Load	CPU
2	2	367	2	[167,200]	0.01
2	4	612	2	[346,266]	0.01
2	6	940	2	[451,489]	0.01
2	8	1189	2	[598,591]	0.03
4	2	318	1	318	0.01
4	4	566	1	566	0.02
4	6	904	1	904	0.02
4	8	1246	1	1246	0.02

Table 29. T=6

I	J	Demand	# of Trucks	Trucks Load	CPU

Table 30. T=8

I	J	Demand	# of Trucks	Trucks Load	CPU

2	2	363	2	[163,200]	0.01	2	2	266	2	[166,100]	0.01
2	4	593	2	[293,300]	0.02	2	4	517	2	[267,250]	0.05
2	6	929	2	[456,473]	0.03	2	6	945	2	[435,510]	0.03
2	8	1207	2	[604,603]	0.04	2	8	1207	2	[700,507]	0.07
4	2	208	1	208	0.01	4	2	280	1	280	0.01
4	4	615	1	615	0.02	4	4	511	1	511	0.02
4	6	902	1	902	0.02	4	6	886	1	886	0.04
4	8	1056	1	1056	0.03	4	8	1218	1	1218	0.04

DISCUSSION

This paper formulated problems P_1 and P_2. One may argue that the problem P_1 does not need to be formulated as a separate transportation problem. One may use a truck with the smallest purchasing cost and assign all transportation to this truck. This approach, however, may not return the best solution as the availability of the budget may help to purchase several trucks and divide transportation among them. Simply choosing trucks with the lowest purchasing cost may not return the best optimum solution. However, problem P_1 returns the optimum number of trucks and their assigned routes.

The numerical analysis section solved and analyzed both problems P_1 and P_2. In problem P_1, as the budget increases, the number of selected trucks remains the same. Because an increase in the budget will not enforce more limitations on the problem.

In problems P_1 and P_2 as the truck capacity increases, the number of selected trucks either remains the same or decreases. Because an increase in capacity will allow a truck to have higher loads. In both problems, as the demand increases, the number of selected trucks either remains the same or increases. Because an increase in demand requires more trucks to fulfill the additional demand.

In problems P_1 and P_2, the number of trucks and their loads are being affected differently when the number of suppliers, number of destinations, and the number of available trucks change. For example, as the number of suppliers increases the number of selected trucks decreases or remains the same. This is not the case when the number of destinations or the number of available trucks decreases. In the latter situation, the number of selected trucks will remain the same. The authors need more numerical analysis to understand the reason behind this observation.

Another observation is that when more than one truck is selected by the solution, the loads of these trucks are quite close to each other. It may be the results of the CBC solver. In the case the solver does not return balanced loads, an appropriate modification to the formulation of the problem is needed. This, however, is beyond the scope of this research and authors recommend additional study on the solver selection.

Throughout the numerical analysis, the transportation cost is assumed to be the same for all trucks. This may not be realistic as quite often the age of trucks and their technology in a fleet may be different which results in a different transportation cost. The authors recommend repeating the numerical study in this paper when the transportation costs are different between trucks.

CONCLUSION AND FUTURE RESEARCH

The transportation problem is extended to account for multiple modes of transportation or multiple time periods. This study, however, extends this problem to account for multiple trucks. The extension is based on two scenarios. In the first scenario, it is assumed a manager has a certain budget and must purchase trucks with a limited capacity. This scenario is formulated as a mixed-integer linear programming problem. The main difference between this formulation and the transportation problem is the presence of trucks. Because of that the focus of the analysis is how the parameters of the formulation affect the number of selected trucks and their loads. Particularly, the numerical analysis section shows the effects of truck capacity, budget, demand, number of suppliers, number of destinations, and the number of available trucks on truck selection and truck loads.

The second scenario assumes that a manager has a fleet of trucks with limited capacity available and must assign them for transportation. This scenario also is formulated as a mixed-integer linear programming problem. The main difference between the second formulation and the transportation problem is the presence of trucks. A numerical analysis is conducted similar to the first scenario. It shows the effects of truck capacity, budget, demand, number of suppliers, number of destinations, and the number of available trucks on truck selection and truck loads.

Three possible future research are recommended. First, the change in the number of destinations and the number of available trucks. The numerical analysis in this paper does not find the reason behind the insensitivity of truck selection to the parameters and more studies are needed. Second, the study shows that loads of trucks remain close and balanced when more than one truck is selected. The reason behind the balanced loads is not clear. It is recommended to use other solvers (this study uses CBC solver) to see if the returned solutions remain balanced or not. If the solutions do not remain balanced using other solvers, then it is recommended to add constraints to enforce a balanced solution. Third, the study assumes the cost of transportation is the same for all trucks. It is recommended to repeat the numerical study in case the transportation costs differ between trucks.

REFERENCES

Balaji, N., & Balaji, A. N. (2018). Solving the multi-period fixed cost transportation problem using LINGO solver. *International Journal of Pure and Applied Mathematics*, 119(12), 2151-2157.

Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (2011). Linear programming and network flows. John Wiley & Sons.

Bookbinder, J. H., & Sethi, S. P. (1980). The dynamic transportation problem: A survey. *Naval Research Logistics Quarterly*, 27(1), 65-87.

Charnsethikul, P., & Svetasreni, S. (2007). The constrained bottleneck transportation problem. *Journal of Mathematics and Statistics*, 3(1), 24-27.

Díaz-Parra, O., Ruiz-Vanoye, J. A., Bernábe Loranca, B., Fuentes-Penna, A., & Barrera-Cámara, R. A. (2014). A survey of transportation problems. *Journal of Applied Mathematics*, 2014.

Flórez, J., de Reyna, A. T. A., García, J., López, C. L., García-Olaya, Á., & Borrajo, D. (2011, March). Planning multi-modal transportation problems. In *Proceedings of the International Conference on Automated Planning and Scheduling* (Vol. 21, pp. 66-73).

Maity, G., Roy, S. K., & Verdegay, J. L. (2020). Analyzing multimodal transportation problem and its application to artificial intelligence. *Neural Computing and Applications*, 32, 2243-2256.

DECISION SCIENCES INSTITUTE

Understanding ERP Continuance Usage from Change Management Perspective

Fan Zhao
Florida Gulf Coast University
Email: fzhao@fgcu.edu

Elias Kirche
Florida Gulf Coast University
Email: ekirche@fgcu.edu

ABSTRACT

Prior research has explored multiple constructs of IS continuance, but the role of change management in ERP continuance usage has been overlooked. We address this gap by integrating change management models and IS continuance theories to establish a foundational research model that examines ERP users' continuance intention in the context of mandatory ERP implementations. We examine mandatory ERP system usage reflecting the realities of the business world, while most ERP continuance studies focus on voluntary usage. Therefore, our model is more accurate to guide ERP implementations projects. We plan to conduct an empirical investigation to validate our research model.

KEYWORDS: Information systems, ERP adoption, Change management, Acceptance model, Continuance intention

INTRODUCTION

Driven by the growth of emerging technologies and rapidly changing innovations in business processes and technological characteristics, organizations are now upgrading their Enterprise Resource Planning (ERP) systems more frequently to adapt to market changes and gain advanced competitiveness. They typically have two options: upgrade the current system with the same vendor or switch to a new ERP system offered by a new vendor. No matter which option they choose, their employees need to accept the new system and continuously use the ERP in their daily business. Therefore, making a smooth transition to the new ERP systems becomes a critical factor influencing the organizations' intention to maximize both short-term and long-term benefits from their ERP investments (Liu et al., 2011; Zhu et al., 2010). However, employees' attitude to continuously use this new ERP system posed a significant challenge. Many employees resisted the new mandatory system by slowing down their productivity, sabotaging, or even refusing to use the system (Malaurent and Avison, 2015).

There have been numerous studies related to continuance intention in various research areas, such as post-implementation of ERP systems (Sun and Mouakket, 2015), social networking sites (Tran et al., 2019), smartphone applications (Tan et al., 2015), e-learning (Aini et al., 2019), and online games (Huang et al., 2018). Although researchers realized continuous usage as one of the critical factors in the post-implementation ERP stage (Cheng, 2018), most of the literature has emphasized voluntary IS usage (Rezvani et al., 2017). However, considering what happens in business practices, policies, and standards, most of the ERP systems are mandatory, which means that employees must use the systems regardless of whether they liked them or not. Therefore, when we study the continuance use of ERP systems, we need to focus more on the continued mandatory system usage, which is more realistic and valuable to the practitioners. Furthermore, even many studies have emphasized the vital role of users in the continuance use of ERP to explain the success of ERP systems, few studies investigate the continuance use of ERP from a change management perspective (Ding, 2019). As one of the critical success factors in ERP implementations (Asemi and Jazi, 2010), change management helps organizations reduce user resistance in ERP adoptions. Unsuccessful change management in post-implementation of ERP may lead to poor decision making and business process performance (Comuzzi and Parhizkar, 2017). Therefore, understanding how change management is involved and could help improve continuance use of mandatory ERP systems is essential for both researchers and organizations.

This study applies change management principles to design a research framework that explains the continuance usage processes of ERP users from denial to acceptance. We address the following research questions: What are key factors associated with change management that enhances the users' continuance intention? How do these factors combine or interact to achieve an ERP post-implementation success? How can we further explore the factors and outcomes of ERP users' continuance usage in future studies? To answer these questions, we propose a foundational research model to show an integrated framework of users' continuance intention and change management under realistic business scenarios where users are using mandatory ERP systems.

This research makes several important theoretical and practical contributions in post-implementation ERP success. First, previous studies have explored technology acceptance and users' continuance intention toward ERP systems, but none has used change management models to understand and enhance users' continuance behaviors in ERP usage, to the best of our knowledge. Change management was discussed mainly on ERP implementation and organizational behavior literature. It is unique to explain individual changes in ERP usage under a change management context. Second, few studies focus on mandatory ERP system usage, which is more closely aligned with the reality that employees typically must use the system whatever they intend to use or not. Mandatory ERP system is different from voluntary systems in most current studies in which users seem to have options to choose their attitude toward the ERP systems. Therefore, studying mandatory ERP system usage can give researchers and practitioners more precise results and valuable suggestions about ERP users' continuance usage. Third, we propose several future studies based on our model. By conducting additional investigations in these areas, researchers can provide further support for practitioners in the development and adoption of ERP systems.

LITERATURE REVIEW

Change Management Models

Change management is one of the primary concerns in organizations that adopted ERP systems (Kim et al., 2005; McAdam and Galloway, 2005). Change management in the ERP context can be interpreted as involving all human, social-related and cultural change techniques needed by management to ease the transition to and minimize organizational resistance of the new ERP environment (Somers and Nelson, 2004). ERP projects bring a massive change to an organization's structure and affect how people work and interact. Therefore, the introduction of a new/upgraded ERP system can cause resistance, confusion, redundancies, and errors. Most of the current change management models are based on the view of changes from organizations' perspectives (Smith and Graetz, 2011). We found two change management models that could be adopted to explain changes from an individual perspective. In the next paragraph we will examine how Lewin's (1947) three-stage model of change and Kubler-Ross's (1969) five-stage model of grief can be integrated to explain employees' perception change in technology acceptance. Both describe several aspects of how people cope with change and loss.

Lewin's (1947) model of change focuses on integrating new techniques for social research and addressing issues in the change within social sciences. At the beginning of the change process, all perceptions are at the freezing stage. Gradually, processes or perceptions start unfreezing, then moving from the old way of doing things to the new one. The last stage, refreezing, involves stabilizing the change by reinforcing the new behaviors and making them part of the new status quo. This model provides a clear outline of the major steps in a change process. However, it has been criticized for failing to recognize the remaining integrated components necessary to fully understand all elements of change (Cummings, 2016). Kubler-Ross (1969) developed a change curve model based on her analysis of the stages of grief that people experience when facing death. This model extended Lewin's model by adding more emotional reactions to change. The change curve model identifies five stages of grief change when a human being faces a relative's death: denial, anger, bargaining, depression, and acceptance (Rosenbaum, 2018). This model was widely recognized in organizational change management (Roose, 2012), technology acceptance research (Sotelo, 2015), and educational studies (Corr, 2019). From a management perspective, the model describes five stages of emotional reactions to change including denial, frustration, depression, experiment, and acceptance (Malone, 2018). During the denial stage, employees refuse the change, similar to the freezing stage in Lew's model. In the next stage, frustration, employees get frustrated by the changes from their previous frozen perceptions. During the third stage, depression, employees' morale, and interests decrease after exploring more aspects of the change and observing more differences in the change from their previous perceptions. This stage is the bottom of the change curve. The next stage, experiment, offers enough opportunities and practices to enhance employees' experience of the change. Therefore, during this stage, employees are getting used to the changes. From Lewin's model, the moving phase interprets these three stages of frustration, depression, and experiment in the change curve of Kubler-Ross. Both models then share the same final stage, acceptance, called refreezing in Lewin's model, which defines the employees' perceptions of intention to accept the change, and will become their new frozen perceptions.

Thus, Lewin's model of change and Kubler-Ross's model of grief can be combined to form an integrated model of perception change that can better explain employees' continuance intention of ERP usage. This model can help managers and researchers understand the psychological processes involved in technology acceptance and design effective interventions to facilitate positive outcomes. In our proposed research model, we use this integrated model to explore and explain users' continuance intention of ERP usage. Table 1 compares the two models and shows how they relate to each other.

IS Acceptance and Continuance Theories

Researchers have applied a variety of acceptance theories and models to explain the factors that impact users' acceptance intention or behaviors toward technology innovations, such as theory of reasoned action (TRA) (Ajzen, 1980), expectation–confirmation model (ECM) (Oliver, 1980), technology acceptance model (TAM) (Davis, 1989), theory of planned behavior (TPB)

(Ajzen, 1991), unified theory of acceptance and use of technology (UTAUT) (Venkatesh, et al., 2003), and diffusion of innovation theory (Rogers, 2010). These theories, models, and extended models have been widely adopted in various acceptance and continuance research. Some theories have been used for identifying the antecedents to continuance intention, such as social cognitive theory (Bandura, 1986), IS continuance theory (Bhattacherjee, 2001), and IS success model (DeLone and McLean, 1992, 2003).

Table 1: Change models

LEWIN'S MODEL	KUBLER-ROSS'S MODEL	DESCRIPTION
Freezing	Denial	Employees resist the change and prefer the status quo
Moving/Unfreezing	Frustration	Employees get frustrated by the change and its implications
Moving	Depressing	Employees lose morale and interest in the change
Moving	Experiment	Employees try out the change and gain experience
Refreezing	Acceptance	Employees accept the change and adopt new behaviors

In this study, we refer to continuance intention as the users' long-term attitude toward a new or updated mandatory ERP system. This paper examines users' behavioral changes over time based on Lewin's model (1947), which encompasses the transition from a state of freezing to refreezing. Additionally, we consider factors associated with each change stage, as defined by Kubler-Ross (1969). These factors include habit, technological characteristics, control, computer anxiety, pressure, perceived usefulness (PU), perceived ease of use (PEOU), and satisfaction. These factors are summarized next and included in our proposed research model shown in Figure 1.

Habit

Limayem et al. (2007) argue that IS users who have been using a particular innovation for a specific time are predisposed to continually use the technology in an automatic and unthinking manner. A user's habit in IS usage becomes inertia to drive his/her continuance intention and less likely to switch to another system. This habit is beneficial to employees' routine jobs with stable processes where the habit turns to automatic steps executed efficiently without awareness (Bargh, 1996). However, when the employees face system changes, habit becomes an obstacle preventing their acceptance and continuance use of a new system. This leads us to posit that ERP users' habits play a significant role in influencing their continued usage of a new or upgraded system.

Technological Characteristics

Sun and Mouakket (2015) argue that considering both system quality and information quality of an enterprise system could better interpret and explain users' continuance intention behaviors. Technological characteristics, including both system quality and information quality, cover all technical aspects of an ERP system, such as flexibility, reliability, integration, accessibility, timeliness, accuracy, completeness, currency, and format. It plays a significant role in determining users' continuance adoption (Limayem and Cheung, 2008). Users' perceptions toward an ERP system mainly rely on the technological characteristics of the system. The user's subjective opinion affects the usefulness (Karahanna & Straub, 1999). Positive outcomes observed by the user will raise the user's attitude about using this technology as well as the intention of using the technology (Subramanian, 1994). Therefore, we posit that the technological characteristics of an ERP system significantly influence users' perceptions of their intention to continue using the system.

Perceived Behavioral Control

Perceived behavioral control refers to how easy or difficult individuals believe it would be for them to perform a behavior (Ajzen, 1991). It is associated with the strength of control belief, which results from the degree that individuals perceive the existence of factors that could inhibit or facilitate the occurrence of the behavior, and power of control belief, which occurs with the power of these factors to make it easier or harder to engage in the behavior (Ajzen, 2002). According to previous research (Ajzen, 1991, Al-Debei et al., 2013), it can be inferred that perceived behavioral control is not only a critical determinant factor influencing users' pre-adoption intentions but also a key indicator affecting users' post-adoption behaviors. Therefore, we hypothesize that the perceived behavioral control associated with an ERP system will significantly influence users' perceptions of their intention to continue using the system.

Computer Anxiety

According to social learning theory (Bandura, 1997), higher anxiety from users' performance experience may lower their efficacy and productivity. Users with computer anxiety perceive fear and apprehension so that they tend to be resistant to and avoidance of computer technology (Heinssen, et al., 1987). In the IS and ERP context, anxiety has been proven to be a personality variable that influences system use (Agarwal and Karahanna, 2000). Numerous studies have revealed that the relationship between anxiety and behavioral intention is mediated by individuals' beliefs (Garaca, 2011), and anxiety is incorporated as an antecedent to ERP users' continuance intention (Chou and Chen, 2009). Users with higher anxiety toward the new or updated ERP systems show less confidence in completing their tasks in the system (Thatcher et al., 2007). Therefore, we hypothesize that computer anxiety will significantly influence users' perceptions of their intention to continue using the system.

Pressure

Social norms refer to accepted societal rules for behavior. Following these rules leads individuals to be accepted in the societal group. It was found that social norms play a significant role in the intention to use any new technology (Hsu and Lu, 2004; Lee, 2009). Previous research has divided the social norm construct into peer influence and supervisor influence to predict users' intention (Cheung and Vogel 2013; Taylor and Todd, 1995). Because one purpose of this research is to examine employees' perceptions of mandatory ERP systems, we attempt to identify the impact of various pressures employees face when they are requested to use a new or updated ERP system. In this study, we focus on peer influence/pressure and managerial influence/pressure on the development of social norms in predicting users' continuance intention. Peer pressure tends to force employees to avoid their peers' disapproval

(Falk and Ichino, 2006), whereas Managerial pressure comes from employees' normative belief of their manager's expectation regarding a specific behavior. Both pressures impact employees perceived subjective norms because employees typically tend to gain their peers' approval and fear the power of their managers to reward or punish their behaviors (Warshaw, 1980). Thus, we believe social norm pressure will be a key factor in the formation of users' continuance intention to a mandatory ERP system.

PU and PEOU

The TAM framework explains a user's attitude towards a technology and the usage effects of the technology. It posits that a user's behavioral intention to adopt a technology is determined by the user's perceived usefulness (PU) and perceived ease of use (PEOU) of the technology (Davis, 1989). Perceived usefulness refers to "the degree to which a person believes that using a particular system would enhance his or her job performance." (Davis, 1989). When the user approaches a new technology, the perception of how this technology enhances performance and productivity influences the user's attitude about the use of this technology (Igbaria et al., 1994). Perceived ease of use refers to "the degree to which a person believes that using a particular system would be free from effort" (Davis, 1989). Ease of use of a technology decreases a user's physical and mental efforts to adapt a new method to complete desired tasks (Saadé & Bahli, 2005). The more the user believes in the ease and effortlessness of a technology, the higher the user's positive perception (Segars & Grover, 1993). Both perceived usefulness and perceived ease of use interpret the logical action behind users' attitude and intention to continuously use a new or updated ERP system.

Satisfaction

Satisfaction refers to a subjective evaluation for a post-usage of a new technology given pre-usage expectations (DeLone and McLean, 1992). Research suggests that the degree of user satisfaction positively affects system usage, which is believed to be a necessary condition for ensuring productivity payoffs from IT investments (Igbaria and Tan, 1997). Therefore, before the benefits of an ERP system can be achieved, end-user satisfaction represents a necessary precondition to facilitate system acceptance and continuance usage. Lee (2010) identifies that satisfaction has the most significant effect on users' continuance intention. Previous studies from the past two decades have shown that satisfaction positively influences behavioral intentions (Azjen, 1980). When end users are satisfied with the upgraded ERP systems, they are more likely to have the behavior intention to keep using the systems. Thus, we posit that users' satisfaction will have a positive influence on their continued usage of a new or upgraded ERP system.

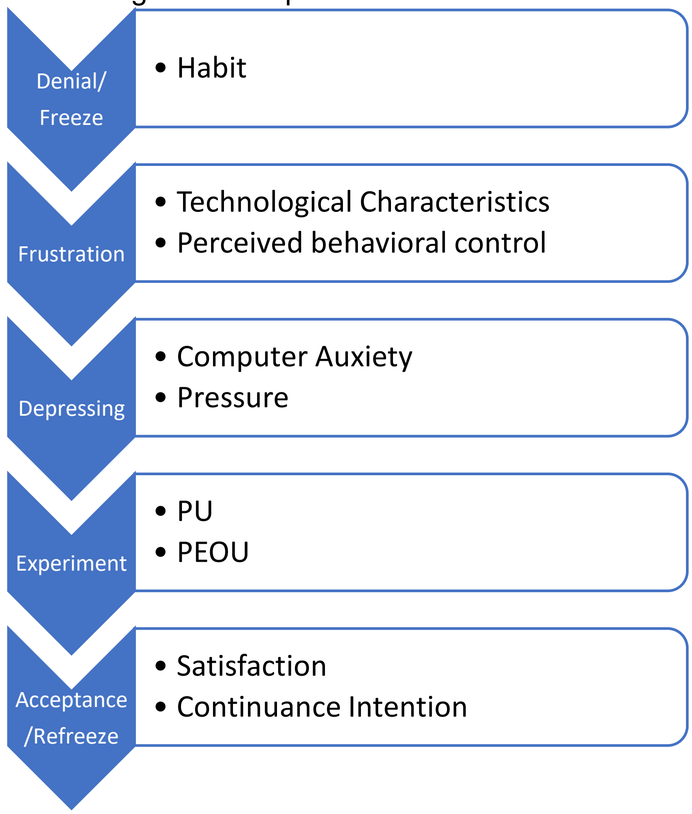
PROPOSED RESEARCH MODEL

Research has shown that effective change management is critical to successfully implement enterprise systems and business process reengineering projects (McAdam and Galloway, 2005; Tsai et al., 2005). A successful organizational change approach relies in a proper integration of people, process, and technology (Esteves & Pastor, 2000). Therefore, we propose a foundational research model to explain users' continuance intention process in ERP usage by combining two change management models and several IS acceptance and continuance theories/models (see Figure 1).

Throughout the change process, Kubler-Ross' change curve model and Lewin's change management model can be aligned to complement each other, especially when one considers the behavioral performance of impacts from changes (Rosenbaum, 2018). We propose the change process to have five stages:

1. Denial/freeze stage: This is the beginning stage that users start to access the new or updated system. In this stage, users' performance habit from the previous system is still solid and frozen. Therefore, users' reaction to the new system is denial, and we adopt the habit construct to explain users' behavior during this stage.
2. Frustration stage: This is when users start to use the new system for either practice or training purposes. Users face a new system including new interfaces, new features and functions, new business processes, etc. Their performance habit starts unfreezing from this stage. Within a brief period of time, users are taking overwhelming information from the system, such as the technological characteristics of the new system, which makes them feel frustrated and lose control of the system with low confidence. In this stage, we believe both technological characteristics of the new system and loss of control will cause users' frustration.
3. Depressing stage: This is the stage at the bottom of the Kubler-Ross change curve. Although users become increasingly familiar with the new system, they perceive the system becoming more complicated and challenging with time after performing more tasks or processes. This is the darkest moment in the learning and practicing process (Rosenbaum et al., 2018). Users' learning experience becomes more stressful, and they begin to show anxiety over their unsuccessful performances in the system. Furthermore, they sense more managerial and peer pressures when they struggle to unfreeze their old habits and form the new habits in the system use. We believe that users' perceptions of anxiety and pressure during this stage bring them depressing feelings toward the new system use.
4. Experiment stage: This is a critical stage influencing users' continuance intention to the new system in the next stage (Rosenbaum et al., 2018). If users perceive both usefulness and ease of use of the new system, they complete their habit unfreezing process and start to build their new habits toward the new system. This will gradually give them the experience of achieving their expectations of using the new system and make them feel more confident about the new system. However, if users perceive low usefulness and ease of use of the new system during this stage, they may face challenges in developing the necessary skills and confidence to fully embrace the system. In such cases, they may experience a return to the Frustration or Depressing stage as they continue to navigate and explore the new system. Typically, the users cannot stop or drop off because the new ERP system is a mandatory system that they must use.
5. Acceptance/Refreeze stage: After users gain a positive experience from the last stage, they tend to be satisfied with the new system, which, in turn, leads to their acceptance of the system and decision to continuously use the new system. After users have a positive experience in the preceding stage, they are more likely to develop satisfaction with the new system, which subsequently influences their acceptance and continuous usage of the system. During this phase, users unconsciously solidify their newly formed habits for the new system. We assert that the constructs of satisfaction and continuance intention are highly relevant in this stage, signifying the completion of the change process.

Figure 1: Proposed Research Model



PROPOSED FUTURE STUDIES

We are planning to conduct an empirical study to test our research model. In 2019, SAP announced that their technical support to all the legacy products, such as ECC and R/3, would stop at the end of 2025. Many companies who run SAP legacy systems are rushing to update their systems. One mid-size organization next to us is updating their SAP R/3 systems to SAP S/4 HANA. Our local government and a large company, with headquarters located in our city, just updated their SAP systems. We will conduct both qualitative (by interviewing CIOs and IT managers in all three organizations) and quantitative (by designing and sending survey questionnaires to both key users and end users in all three organizations) studies to empirically test our research model and answer the following research questions:

- Does this combination of change management and continuance theories effectively explain users' resistance and enhance their intention to continue using ERP systems?
- Do the current constructs in our research model fit in each change management stage?
- Is each of the current constructs in our research model an indicator affecting users' post-implementation behaviors?
- Are there any additional factors or constructs that should be included in this model?
- How do end users feel about the change stages?
- Are there specific change stages that occur in ERP upgrade projects?
- What is the typical duration for ERP users to unfreeze and refreeze their habits?

CONCLUSIONS

Numerous studies have explored the antecedents of ERP users' continuance behavior during the post-implementation stage. However, the role of change management in fostering users' continuance intention towards mandatory ERP systems has received little or no attention. This study addresses this gap by integrating two change management models and several IS continuance theories to establish a research model that examines ERP users' continuance intention from a change management perspective. Our research model defines five change management stages, each incorporating one or two IS continuance constructs: the Denial/Freeze stage with habit, the Frustration stage with technological characteristics and perceived behavioral control, the Depression stage with computer anxiety and pressure, the Experiment stage with perceived usefulness and perceived ease of use, and the Acceptance/Refreeze stage with satisfaction and continuance intention to the system.

Our study has two theoretical implications. Firstly, prior research (Bhattacherjee, 2001; Chou and Chen, 2009) has explored multiple constructs of IS continuance, but the role of change management in ERP continuance usage has been overlooked. Our study uniquely integrates theories and models from both IS continuance frameworks and change management models. This research model expands the understanding of IS continuance, particularly in the context of ERP systems. Secondly, while most existing ERP continuance studies focus on voluntary system usage, we specifically examine mandatory ERP system usage, which closely reflects the realities of the business world where employees are required to use the adopted ERP system. From a practitioner's perspective, applying current ERP continuance studies directly to their ERP projects can be challenging as many results are too conceptual and may not account for mandatory ERP systems. Therefore, our model and findings can be readily applied to ERP project management as we combine the change management perspective with IS continuance theories and emphasize the usage of mandatory ERP systems in our model.

REFERENCES

Agarwal, R. and Karahanna, E. 2000. "Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage," *MIS Quarterly (*24:4), pp. 665-694.

Aini, Q., Rahardja, U., and Hariguna, T. 2019. "The antecedent of perceived value to determine of student continuance intention and student participate adoption of ilearning," *Procedia Computer Science* (161), pp. 242-249.

Azjen, I. 1980. *Understanding attitudes and predicting social behavior*, *Englewood Cliffs*.

Ajzen, I. 1991. "The theory of planned behavior," *Organizational Behavior and Human Decision Processes (50:*2), pp. 179-211.

Ajzen, I. 2002. "Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior," *Journal of applied social psychology* (*32*:4), pp. 665-683.

Al-Debei, M., Al-Lozi, E., and Papazafeiropoulou, A. 2013. "Why people keep coming back to Facebook: Explaining and predicting continuance participation from an extended theory of planned behavior perspective," *Decision support systems* (55:1), pp. 43-54.

Asemi, A., and Jazi, D. 2010. "A comparative study of critical success factors (CSFs) in implementation of ERP in developed and developing countries," *International Journal of Advancements in Computing Technology* (*2:*5), pp. 99-110.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Cognitive Social Theory*, Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. 1997. *Self-Efficacy: The exercise of Control*, W.H. Freeman, New York.

Bargh, J. 1996. "Automaticity in Social Psychology," in *Socialpsychology: Handbook of Basic Principles,* Higgins ET and Kruglanski AW (eds.), Guilford, New York, pp. 169-183.

Beckers, J. and Schmidt, H. 2003. "Computer experience and computer anxiety," *Computers in Human Behavior* (19:6), pp. 785-797.

Bhattacherjee, A. 2001. "Understanding information systems continuance: An expectation-confirmation model," *MIS quarterly* (25:3), pp. 351-370.

Cheng, Y. 2018. "What drives cloud ERP continuance? An integrated view," *Journal of Enterprise Information Management*, (31:5), pp. 724-750.

Cheung, R., & Vogel, D. 2013. "Predicting user acceptance of collaborative technologies: An extension of the technology acceptance model for e-learning," *Computers & Education* (63), pp. 160–175.

Chou, S. and Chen, P. 2009. "The influence of individual differences on continuance intentions of enterprise resource planning (ERP)," *International Journal of Human-Computer Studies* (67:6), pp. 484-496.

Comuzzi, M., and Parhizkar, M. 2017. "A methodology for enterprise systems post-implementation change management," *Industrial Management & Data Systems* (117:10), pp. 2241-2262.

Corr, C. 2019. "Should We Incorporate the Work of Elisabeth Kübler-Ross in Our Current Teaching and Practice and, If So, How?" *OMEGA-Journal of Death and Dying* (0:0), pp. 1-23.

Cummings, S., Bridgman, T., & Brown, K. 2016. "Unfreezing change as three steps: Rethinking Kurt Lewin’s legacy for change management," *Human relations* (*69:*1), pp. 33-60.

Davis, F. 1989. "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly* (13:3), pp. 319–340.

DeLone, W. and McLean, E. 1992. "Information systems success: the quest for the dependent variable," *Information Systems Research* (3:1), pp. 60–95.

DeLone, W. and McLean, E. 2003. "The DeLone and McLean model of information systems success: a ten-year update," *Journal of management information systems* (*19:*4), pp. 9-30.

Ding, Y. 2019. "Looking forward: The role of hope in information system continuance," *Computers in Human Behavior* (91), pp. 127-137.

Falk, A. and Ichino, A. 2006. "Clean evidence on peer effects," *J. of Labor Economics* (24), pp. 39–58.

McAdam, R., & Galloway, A. 2005. "Enterprise resource planning and organizational innovation: a management perspective, " *Industrial Management & Data Systems* 105(3), pp. 280-290.

Oliver, R. 1980. "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of marketing research* (*17:*4), pp. 460-469.

Rezvani, A., Dong, L., and Khosravi, P. 2017. "Promoting the continuing usage of strategic information systems: The role of supervisory leadership in the successful implementation of enterprise systems," *International Journal of Information Management* (37:5), pp. 417-430.

Rogers, E. 2010. *Diffusion of innovations* (4th ed.), New York: The Free Press.

Roos, S. 2012. "The Kubler-Ross model: an esteemed relic," *Gestalt Review* (*16:*3), pp. 312-315.

Rosenbaum, D., More, E., and Steane, P. 2018. "Planned organizational change management: Forward to the past? An exploratory literature review," *Journal of Organizational Change Management (31:2), pp. 286-303*.

Somers, T., & Nelson, K. 2004. "A taxonomy of players and activities across the ERP project life cycle," *Information & Management* (41:3), pp. 257-278.

Smith, A. and Graetz, F. 2011. *Philosophies of Organizational Change*, Edward Elgar Publishing.

Sotelo, B. 2015. "*A qualitative case study to investigate the technology acceptance experience outlined in the TAM using the Kübler-Ross stages of grieving and acceptance,*" (Doctoral dissertation, Colorado Technical University).

Sun, Y., and Mouakket, S. 2015. "Assessing the impact of enterprise systems technological characteristics on user continuance behavior: An empirical study in China," *Computers in Industry* (70), pp. 153-167.

Tan, W., Lee, P., and Hsu, C. 2015. "Investigation of temporal dissociation and focused immersion as moderators of satisfaction–continuance intention relationship: Smartphone as an example," *Telematics and Informatics* (32:4), pp. 745-754.

Taylor, S., and Todd, P. 1995. "Understanding information technology usage: A test of competing models," *Information Systems Research* (6:2), pp. 144–176.

Thatcher, J., Loughry, M., Lim, J., and McKnight, D. 2007. "Internet anxiety: An empirical study of the effects of personality, beliefs, and social support," *Information & Management* (44:4), pp. 353-363.

Tran, L., Pham, L., and Le, L. 2019. "E-satisfaction and continuance intention: The moderator role of online ratings," *International Journal of Hospitality Management* (77), pp. 311-322.

Tsai, W., Chien, S., Hsu, P., & Leu, J. 2005. "Identification of critical failure factors in the implementation of enterprise resource planning (ERP) system in Taiwan's industries," *International Journal of Management and Enterprise Development* (2:2), pp. 219-239.

Venkatesh, V., Morris, M., Davis, G., and Davis, F. 2003. "User acceptance of information technology: Toward a unified view," *MIS quarterly* (27:3), pp. 425-478.

Warshaw, P. 1980. "A new model for predicting behavioral intentions: An alternative to Fishbein," *Journal of marketing research* (17:2), pp. 153-172.

Zhu, Y., Li, Y., Wang, W., and Chen, J. 2010. "What leads to post-implementation success of ERP? An empirical study of the Chinese retail industry," *International Journal of Information Management* (30:3), pp. 265-276.

DECISION SCIENCES INSTITUTE

Understanding Individuals' Internet Security Threat Ambivalence, Approach, and Avoidance

Yaojie Li
University of New Orleans
Email: yli27@uno.edu

Xuan Wang
University of Texas Rio Grande Valley
Email: xuan.wang@utrgv.edu

Hanieh Javadi Khasraghi
University of Delaware
Email: hanieh@udel.edu

ABSTRACT

In the prior security literature, there is limited understanding of individual Internet security behavior in non-work circumstance. Hence, this paper aims to explain how individuals' threat and coping appraisals affect their different behavioral outcomes, i.e., threat avoidance or threat approach. Our findings demonstrate that, consistent with prior studies, threat avoidance is influenced by an individual's perceived threat, response efficacy, and response cost, whereas threat approach is influenced only by individuals' perceived response efficacy and response cost. Also, our study reveals that threat ambivalence is a significant mediator in the associations between response efficacy, response cost, and threat approach.

KEYWORDS: Information security, Internet security, Threat avoidance, Threat approach, and Threat ambivalence

INTRODUCTION

The Internet is an integral part of our lives today and will be even more influential as we return to a new normal after the pandemic (McClain et al., 2021). According to Broadband Search (2022) – an online database of all the Internet and TV providers in the U.S., approximately 5.25 out of the entire 7.9 billion people worldwide have access to and use the Internet frequently. Also, Internet usage has rapidly surged during the pandemic on social media, video platforms, and eBusiness. Nevertheless, while individuals are singing praise for the benefits and opportunities the Internet brings, individuals' security and privacy concerns simultaneously overshadow their confidence about online activities. To resolve this dilemma, many information security practitioners and theorists have made substantial progress in providing best practices and theoretical directions about safe and secure online behaviors.

Among a large number of Internet security behavioral research, two current streams of thought are more prevalent than others. One is the protection motivation theory (PMT), which aims to predict individuals' intentions to protect themselves after receiving fear-arousing recommendations (Floyed et al., 2000). In this sense, an individual who receives a security fear inducing message will likely engage in an adaptive response through threat appraisal and coping appraisal, ultimately taking protective actions (cf. Anderson & Agarwal, 2010; Boss et al., 2015; Posey et al., 2015). Another framework used in individual Internet usage is the technology threat avoidance theory (TTAT). Drawing on behavioral control theories (Carver et al., 1982; Edwards, 1992), Liang and Xue (2009) proposed a cybernetic process in which individuals intend to continually enlarge the distance between their current security state and the undesired end state through threat appraisal and coping appraisal.

While both PMT and TTAT accentuate the importance of threat appraisal and coping appraisal when individuals cognitively engage in dealing with security threats, their predicted outcomes are mainly positive security behaviors (i.e., protection or threat avoidance). However, few studies have investigated how individuals approach security threats, an oppositional outcome that can coincide with individuals' threat avoidance or protective action. Those threat approaching activities can be attributed to individuals' careless and hasty clicks, curiosity to explore new things online, or pursuit of shortsighted benefits from "free" yet unsafe websites.

Paradoxically, while TTAT points out that the approach-avoidance distinction was often ignored, current research has rarely examined threat approach. One explanation is that the original assumption of TTAT is based on the difference between IT adoption and threat avoidance, without considering specific contexts where there are no such physical systems and software adoption in organizations but websites and Internet platforms at home. Another reason is that, in addition to social desirability bias toward negative security behaviors, valid self-report data on deviance or omission via conventional forms of survey research is relatively difficult to obtain.

Revisiting the threat approach-avoidance distinction, we noticed a research deficiency in explicating the intermediary construct between threat and coping appraisals and contradictory results when individuals betray due to mixed perceptions and thoughts about Internet security threats or threat ambivalence. Inspired by emerging information systems (IS) research on mixed mindsets and ambivalence (Jarvenpaa & Majchrzak, 2010; Lapointe & Beaudry, 2014; Qahri-Saremi & Turel, 2020; Stein et al., 2015), it is imperative to extend the landscape of PMT and TTAT by incorporating ambivalence and exploring its interactions with threat approach and avoidance. Addressing this research gap is paramount for at least two reasons. First, various Internet security threats and individual differences (e.g., risk tolerance, security capabilities) can lead to different personal experiences and perspectives that provide living space for mixed security attitudes and emotions. To our best knowledge, this cognitive and attitudinal phase has rarely been examined in IS security research. Undiscovered mediators such as threat ambivalence can unravel the black box between the appraisal processes and coping behaviors while explaining the inconsistent and contradictory results found in previous TTAT studies (Carpenter et al., 2019). In other words, introducing threat ambivalence can better explain how and why individuals engage in oppositional security behaviors – threat avoidance and threat approach. Our study can thus contribute to Internet security behavioral research in various ways previously overlooked.

In this paper, we strive to address the research gap by investigating the role of threat ambivalence in coping appraisal and threat appraisal as individuals encounter Internet threats and how this ambivalence status will affect their consequent security behaviors. Hence, our research objectives are manifested in two main questions: 1) how does threat ambivalence emerge from threat appraisal and coping appraisal? And 2) how does the ambivalence influence individuals' Internet threat approach and avoidance?

LITERATURE REVIEW

Individual user security has become an important topic in IS security research since consumer devices have rapidly increased and expanded information security spheres, and hackers can hijack those personal devices to threaten organizational information security (Li & Siponen, 2011). Information security behaviors at the workplace are usually subject to formal, mandatory control measures, such as organizational information security policies and procedures, technical monitoring, security education, training, and awareness (SETA) programs. Therefore, deterrence mechanisms built on the severity, certainty, and celerity of punishments toward security misbehavior such as policy violations have dominated IS security research for many years (e.g., Chen et al., 2012; D'Arcy et al., 2009; Herath & Rao, 2009; Straub & Welke, 1998). Unlike security behaviors in the formal, working context, security behaviors in non-work circumstances are more voluntary and subjective, and individuals can choose whether and how they engage in security activities. Further, organizational employees can benefit from security training and support provided by their technical colleagues (i.e., established security countermeasures against threats), whereas most individuals in the home context have to confront and cope with security threats by themselves. For that, theories related to individual affective and cognitive security engagement, such as the theory of planned behavior, protection motivation theory, and technology threat avoidance theory, seem more appropriate in an individual, non-work security scenario. Hence, the research topics in the non-work context tend to be more specific, such as avoiding phishing emails, unsafe websites, and malware, compared with organization-oriented themes such as security policy compliance or violations (e.g., Bulgurcu et al., 2010; Moody et al., 2018; Puhakainen & Siponen, 2010; Siponen & Vance, 2010). Specifically, technology threat avoidance theory is an increasingly popular perspective to explain and predict individual Internet security behavior in the non-work context. Furthermore, as discussed earlier, security behaviors in the non-work context appear more voluntary and subjective than those in organizations. Hence, individuals' self-control mechanisms featured in the technology threat avoidance theory become more salient than organizational security controls such as security sanctions, rewards, norms, and cultures.

Technology threat avoidance theory (TTAT) is an indigenous information systems security theory developed by Liang and Xue (2009). They synthesized cybernetic and coping theories that are prominent in the fields of psychology, healthcare, and risk analysis. TTAT delineates the avoidance behavior as a dynamic, positive feedback loop in which individuals experience two cognitive processes, threat appraisal and coping appraisal, ultimately deciding how to cope with IT threats. Also, TTAT suggests that individuals would set up the anti-goal (i.e., being harmed by malicious IT) and perform threat appraisal and coping appraisal before engaging in emotion-focused and problem-focused coping activities. Liang and Xue (2010) further examined the process theory in a follow-up empirical study (2010). Two constructs – perceived threat and perceived avoidability – constitute the main antecedents of coping behaviors, where their perceived susceptibility and severity determine individuals' perceived threat. Meanwhile, individual users' perceived avoidability is contingent upon the perceived effectiveness and costs of the security measure and the self-efficacy of applying the security measure. The avoidance process will continue to evolve as new malicious IT threats emerge and evolve. While TTAT proposed the cybernetic loop based on how individuals assess and avoid malicious IT, it seems to ignore that individuals may insist on approaching Internet security threats (e.g., visiting unsafe websites and downloading suspicious malware), even acknowledging the threats. In other words, the approach-avoidance dynamics were partially reflected in the TTAT model. To that end, it is necessary to enhance the internal validity and expand the theoretical boundary of TTAT by examining both threat avoidance and approach while considering the transient construct – threat ambivalence.

Ambivalence is a state of mind that occurs when individual experiences simultaneous positive and negative orientations toward an object, and this term can be cognitive or emotional, functional or dysfunctional (Ashforth et al., 2014). IS studies are turning increasingly to this construct and linking it to IT usage (Bala et al., 2017; Lapointe & Beaudry, 2014; Qahri-Saremi & Turel, 2020; Stein et al., 2015), online community and knowledge collaboration (Jarvenpaa & Majchrzak, 2010; Ludwig et al., 2014), e-commerce and online-transactions (Moody et al., 2014; 2017), and information security (Ng et al., 2021). However, the only application of the ambivalence construct in the behavioral security area is Ng et al. (2021). Ng and colleagues (2021) focus on the ambivalence based on the opposition between individuals' evaluations of maladaptive rewards and social norms, ultimately affecting individuals' protection behavior.

Instead, our study assumes that individuals can have simultaneous conflicting reactions toward security threats while assessing security threats and related response measures.

According to Liang and Xue (2009), perceived threat refers to the proximity between an individual's current and undesired end states when perceiving malicious IT as dangerous. Motivated by extant psychology and risk analysis studies, Liang and Xue (2009) proposed two antecedents – perceived susceptibility and perceived severity – to perceived threat. Perceived susceptibility refers to individual users' belief about the degree of vulnerability to malicious IT, whereas perceived severity is individual users' belief about the significance or magnitude of potential harm caused by malicious IT (Liang & Xue, 2009). We, too, argue that individuals can confront similar security threats while visiting unsafe websites or engaging in another Internet security activities in a non-work context (e.g., at home or public areas). Likewise, individuals can better perceive Internet security threats when they believe their computers and devices are vulnerable to malware and attacks from the Internet. Furthermore, the severity of negative consequences caused by security threats will shape their awareness and understanding of the threats. Hence, we hypothesize that:

Hypothesis 1a: Individual users' perceived susceptibility is positively associated with their perceived threat.

Hypothesis 1b: Individual users' perceived severity is positively associated with their perceived threat.

The pleasure-unpleasure principle illuminates the human's instinctive pursuit of pleasure and avoidance of pain to satisfy biological and psychological needs (Freud, 1958). Likely, when individuals face Internet security threats that can cause privacy disclosure and financial losses, they are motivated to cease approaching or avoid security threats. Indeed, the positive associations between perceived security threats and preventative and protective behaviors have been widely supported in previous behavioral security studies (e.g., Arachchilage & Love, 2013; Boss et al., 2015; Chen & Zahedi, 2016; Liang & Xue, 2010; Posey et al., 2015). However, due to the bounded rationality in assessing the threats and related coping measures, an individual can also feel ambivalent. Moreover, their felt ambivalence can increase as the perceived threat rises. Hence, we posit:

Hypothesis 2a: Individual users' perceived threat is positively associated with their threat avoidance.

Hypothesis 2b: Individual users' perceived threat is negatively associated with their threat approach.

Hypothesis 2c: Individual users' perceived threat is positively associated with their threat ambivalence.

The threat coping appraisal appears in various theoretical frameworks, including protection motivation theory and the theory of planned behavior (e.g., Burns et al., 2017; Lee & Larsen, 2009; Liang et al., 2019; Grimes & Marquardson, 2019). Individuals who have perceived security threats will enter the following phrase of the avoidance route – coping appraisal (Liang & Xue, 2009). In this process, individual users would evaluate the efficacy of related security countermeasures against Internet threats, i.e., response efficacy. When an individual has low level confidence in the security countermeasures, he or she should be very cautious about the threat and avoid it at most. However, when individuals are confident about their security response measures (e.g., updated anti-virus software and authenticated networks), they may enter a dilemma between avoidance and approach. They may even take further risks and approach Internet security threats (e.g., visiting unsafe websites). Hence:

Hypothesis 3a: Individual users' perceived security response efficacy is positively associated with their threat avoidance.

Hypothesis 3b: Individual users' perceived security response efficacy is negatively associated with their threat approach.

Hypothesis 3c: Individual users' perceived security response efficacy is negatively associated with their threat ambivalence.

Response costs refer to any perceived direct personal cost (e.g., effort, time, money, or trouble) incurred when individuals cope with security threats (Floyd et al., 2000; Liang & Xue, 2010). Hence, individuals will evaluate the costs when making relevant coping choices. It can be assumed that when the response costs are reasonable (e.g., price-reasonable and easy-to-use protective software), most individuals are likely to implement it. However, when the response costs are relatively high or more than the benefits obtained from approaching risky Internet sources, individuals may feel ambivalent and choose to approach threats. Thus,

Hypothesis 4a: Individual users' perceived response cost is negatively associated with their threat avoidance.

Hypothesis 4b: Individual users' perceived response cost is positively associated with their threat approach.

Hypothesis 4c: Individual users' perceived response cost is positively associated with their threat ambivalence.

Ambivalence is likely to trigger individual biased systematic information processing, which will lead to a closer correspondence between attitudes and intentions (Jonas et al., 1997; Ng et al., 2021). In the scenario of Internet security, individuals will follow their prior ambivalent status (when they realize the threat is acceptable after assessing it and relevant protective measures) and approach security threats. Therefore:

Hypothesis 5a: Individual users' threat ambivalence is negatively associated with their threat avoidance.

Hypothesis 5b: Individual users' threat ambivalence is positively associated with their threat approach.

RESEARCH METHOD

Survey Setting

We collected our data through a field survey from three American universities of various sizes (approximately 6,000 to 30,000 students) and settings (e.g., teaching vs. research). The participants were mainly a mixture of undergraduate and graduate, traditional and adult students who participated in exchange for extra course credit. We illustrate the demographics and IT relevant descriptive statistics in Table 1. In sum, we received 294 responses, of which 218 valid responses (74%) remained after filtering out inattentive and incomplete responses and extreme outliers.

Table 1: Demographics of Respondents

Gender	Age	Education	Work experience	IT proficiency
Male: 125 (57.34%)	20 & below: 100 (45.87%)	Some coll. credits: 162 (74.31%)	Range: 0-30	Fundamental: 46 (20.10%)
Female: 93 (42.66%)	21-30: 105 (48.16%)	Assoc. degree: 24 (11.01%)	Mean: 3.75	Novice: 60 (27.52%)
	31 & above: 13 (5.97%)	Bachelor's degree: 25 (11.47%)	Std. dev.: 4.35	Intermediate: 91 (41.74%)
		Master's degree: 2 (0.92%)		Advanced: 19 (8.72%)
		Doc. Degree: 5 (2.29%)		Expert: 2 (0.92%)
Total: 218 (100%)				

Measurement Development

The measurement items were adapted from the well-established studies in the prior literature and developed from a focus group of Internet users consisting of university faculty and students (see measurement items in Appendix). We performed a preliminary analysis to assess basic psychometric properties and retain the most reliable measures for fitting the hypothesized model. Lastly, we refined eight constructs for further analysis, including perceived susceptibility with four indicators, perceived severity with five indicators, perceived threat with five indicators, threat avoidance with seven indicators, threat ambivalence with five indicators, threat approach with six indicators, response efficacy with four indicators, and response cost with five indicators.

Exploratory Factor Analysis

We conducted an exploratory factor analysis (EFA) to explore the factor structure of constructs while reducing cross-loading items. Principal axis analysis with varimax rotation was applied to identify variables highly associated with the model's constructs. After that, we identified 33 items characterized by factor loadings above a threshold value of 0.4. Table 2 illustrates excessive consistency among the items under each factor with their respective factor-loadings. The exogenous constructs of the study were perceived threat, response efficacy, and response cost; the endogenous constructs were various assessments of Internet threat ambivalence, avoidance, and approach.

Table 2: Finalized Indicator Loadings

	Component				
	1	2	3	4	5
Perceived Susceptibility	0.828	0.786	0.733	0.700	
Perceived Severity	0.792	0.791	0.713	0.657	0.540
Perceived Threat	0.716	0.714	0.635	0.603	
Response Efficacy	0.849	0.844	0.812	0.796	
Response Cost	0.753	0.748	0.626	0.610	0.572
Threat Avoidance	0.812	0.667	0.551		
Threat Ambivalence	0.931	0.882	0.751	0.652	
Threat Approach	0.709	0.688	0.660	0.544	

Reliability and Validity

Reliability and validity are the primary focus of structural equational modeling studies (Hair et al., 2006). According to Nunnally (1994), reliability levels beyond 0.7 form a threshold to ensure that results are reasonably free of measurement error and perform in a reliable manner. In our analysis, construct reliability scores across the overall study exceed 0.7 (Table 3). However, we also assessed reliability as part of investigating the trait validity features of convergence and discrimination in our construct validation process (Boudreau et al., 2001; Henseler et al., 2015; MacKenzie et a., 2011). Also, the research model fits the data well, as all the composite reliability scores on constructs and Cronbach's alphas scores for individual scales are greater than 0.7. Furthermore, the average variances extracted (AVE) values are larger than the square of the individual correlations among constructs. Therefore, we obtain sound evidence supporting convergent and discriminant validity among the reflective constructs in the model.

First, we performed Harman's (1976) single factor test – with the first factor explaining less than the 50% threshold of the variance, indicating that no single factor contributed to the majority of the variance (Podsakoff et al., 2003). Second, we employed a full collinearity assessment approach for PLS-based SEM (Hair et al., 2006; Kock, 2015) – each construct was placed as the outcome variable to test the variation inflation factor (VIF), and all the VIF values obtained (all less than 1.5) are less than the threshold of 3.3. Hence, we conclude that common method bias is not significant in this study.

Table 3: Construct Reliability and Validity

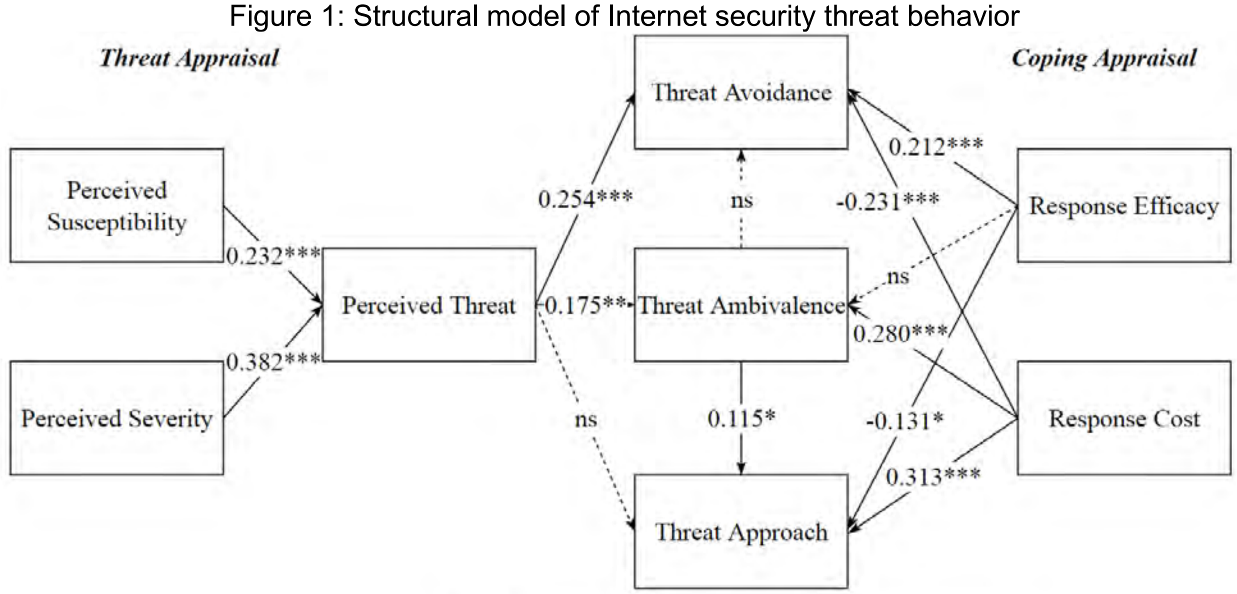
	Cronbach's Alpha	Composite Reliability	AVE
Perceived Susceptibility	0.876	0.915	0.729
Perceived Severity	0.846	0.891	0.622
Perceived Threat	0.815	0.878	0.644

Response Efficacy	0.789	0.862	0.610
Response Cost	0.789	0.862	0.610
Threat Avoidance	0.788	0.876	0.703
Threat Ambivalence	0.893	0.924	0.752
Threat Approach	0.789	0.862	0.610

ANALYSIS AND RESULTS

In this study, we used the PLS-SEM assessment of the path model due to its advantage in theoretical prediction in complex models (Hair et al., 2006). We also implemented a standard bootstrap resampling procedure (5,000 samples) to test the path significance. As shown in Figure 1, significant effects, supportive of hypothesized expectations, are found for all paths except the relationships between perceived threat and threat approach and response efficacy and threat ambivalence. The threat appraisal side (left side, Figure 1) shows that when an individual perceives the vulnerability of their digital devices and the significance of potential harm caused by Internet security threats, their concerns about the threats will correspondingly increase. Also, as one perceives Internet security threats, they will not approach the threats such as visiting an unsafe website, while their intention to avoid the threat will be increased. While there is no direct association between perceived threat and threat approach, an individual will likely approach Internet security threats through a transitional process of feeling ambivalence. In other words, individuals can immediately avoid the security risk once they perceive threats. However, if they start to feel ambivalent about the threats, they will likely choose to take the risk. The coping appraisal (right side, Figure 1) reveals that individual perceived response efficacy is positively associated with threat avoidance and negatively associated with the threat approach. There is no space for ambivalence or hesitation, which appears to be distinct individual security decision-making. When individuals perceive high response costs, their intention to avoid threats will be diminished. Instead, they may experience ambivalence and then choose to approach Internet security threats for some purposes, such as free Internet resources, or simply satisfy their curiosity.

Figure 1: Structural model of Internet security threat behavior



Note. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ns = insignificant

DISCUSSION AND IMPLICATIONS

Our results support our assumption about the different directions of individual security behaviors after completing threat and coping and appraisals, i.e., threat approach and threat avoidance. As technology threat avoidance theory (TTAT) suggests, we found the important roles of perceived threats, response efficacy, and response costs in determining individuals' threat avoidance behavior. Moreover, in our extended TTAT model, we shed light on how the appraisal antecedents affect individuals' threat approach when financial or personal interests drive them. Clearly, individuals' perceived threat cannot directly prevent their threat approach behavior, yet their assessment of the security response mechanism and costs will intervene in their approaching Internet threats. Furthermore, our study examines the mediating role of threat ambivalence, a transient construct, between individuals' cognitive processes and behavioral outcomes. First, we obtained significant results from a full mediation model based on the perceived threat, ambivalence, and threat approach. Second, we found significant results from a partial mediation model based on response cost, threat ambivalence, and threat approach.

Despite important results supporting the extended TTAT model for individual security behavior in the non-working context, several limitations exist and provide possible directions for future research. First, this study is limited in its potential generality issue by using a convenience study sample, though our data include many students with abundant work experience and adult students who are full-time employees. Besides the non-work context, we speculate that a different sample using entire full-time company employees may lead to intriguing results that can be consistent or contradictory to our findings. Since we found that individuals can take risks and approach Internet security threats though they have experienced the threat and coping appraisal, it is worthwhile to investigate how and why they will be motivated to take the risk, for example, financial interest, individual curiosity, or bounded rationality in assessing the risk and coping. Lastly, confirmatory support and novel insights can be generated through different research methodologies (e.g., field experiments, qualitative interviews) and theoretical lenses (e.g., balance theory, rational-choice theory, neural-IS theories).

Besides the opportunities for future research, the practical implications of our work here are evident. First, our findings on the possibility of a threat approach call for attention to individual cognitive abilities in the threat and coping appraisals. In that sense, it is paramount to enhance the efficiency of relevant security education and training programs for individuals at the workplace and non-workplace. Second, threat ambivalence is vital in inducing individuals to the "dangerous" side of the Internet, based on our findings. Hence, it is necessary to employ organizational and psychological methods to mitigate the effects of threat ambivalence and strengthen individuals' determination to avoid Internet security threats and risks.

CONCLUSIONS

Individual security in the non-work context is equally important in the workplace nowadays. However, in contrast to formal and mandatory security control implemented in the organizational setting, individuals often lack sufficient guidance and resources to secure their digital devices and networks and are more likely to approach Internet security threats. Hence, it is necessary to conduct more thorough research to understand individuals' threat assessment, coping, ambivalence, and consequent security behavioral choices and furnish their arsenal in coping with numerous Internet security threats.

REFERENCES

Anderson, C. L., & Agarwal, R. (2010). Practicing safe computing: A multimethod empirical examination of home computer user security behavioral intentions. MIS Quarterly, 34(3), 613-643.

Arachchilage, N. A. G., & Love, S. (2013). A game design framework for avoiding phishing attacks. Computers in Human Behavior, 29(3), 706-714.

Ashforth, B. E., Rogers, K. M., Pratt, M. G., & Pradies, C. (2014). Ambivalence in organizations: A multilevel approach. Organization Science, 25(5), 1453-1478.

Bala, H., Labonté-LeMoyne, E., & Léger, P. 2017. Neural Correlates of Technological

Ambivalence: A Research Proposal. In Information systems and neuroscience, Davis, F., Riedl, R., Brocke, J., Léger, P. & Randolph, A. (Eds., pp. 83-89). New York, NY: Springer.

Boss, S. R., Galletta, D. F., Lowry, P. B., Moody, G. D., & Polak, P. (2015). What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. MIS Quarterly, 39(4), 837-864.

Boysen, S., Hewitt, B., Gibbs, D., & McLeod, A. (2019). Refining the threat calculus of technology threat avoidance theory. Communications of the Association for Information Systems, 45(1), 95-115.

Boudreau, M. C., Gefen, D., & Straub, D. W. (2001). Validation in information systems research: A state-of-the-art assessment. MIS Quarterly, 25(1), 1-16.

Broadband Search. (2022). Key Internet statistics to know in 2022 (Including mobile). Available from: https://www.broadbandsearch.net/blog/internet-statistics

Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Quarterly, 34(3), 523-548.

Burns, A. J., Posey, C., Roberts, T. L., & Lowry, P. B. (2017). Examining the relationship of organizational insiders' psychological capital with information security threat and coping appraisals. Computers in Human Behavior, 68, 190-209.

Carpenter, D., Young, D. K., Barrett, P., & McLeod, A. J. (2019). Refining technology threat avoidance theory. Communications of the Association for Information Systems, 44(1), 380-407.

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. Psychological Bulletin, 92(1), 111-135.

Chen, H., & Li, W. (2017). Mobile device users' privacy security assurance behavior: A technology threat avoidance perspective. Information & Computer Security, 25(3), 330-344.

Chen, Y., Ramamurthy, K., & Wen, K. W. (2012). Organizations' information security policy compliance: Stick or carrot approach? Journal of Management Information Systems, 29(3), 157-

188.

Chen, Y., & Zahedi, F. M. (2016). Individuals' Internet Security Perception and Behaviors: Polycontextural Contrasts between the United States and China. MIS Quarterly, 40(1), 205-222.

D'Arcy, J., Hovav, A., & Galletta, D. (2009). User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. Information Systems Research, 20(1), 79-98.

Edwards, J. R. (1992). A cybernetic theory of stress, coping, and well-being in organizations. Academy of Management Review, 17(2), 238-274.

Floyd, D. L., Prentice-Dunn, S., & Rogers, R. W. (2000). A meta-analysis of research on protection motivation theory. Journal of Applied Social Psychology, 30(2), 407-429.

Freud, S. (1958). Formulations on the two principles of mental functioning. In The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XII (1911-1913): The Case of Schreber, Papers on Technique and Other Works (pp. 213-226).

Grimes, M., & Marquardson, J. (2019). Quality matters: Evoking subjective norms and coping appraisals by system design to increase security intentions. Decision Support Systems, 119, 23-34.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). Multivariate data analysis (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice-Hall.

Harman, H. H. (1976). Modern Factor Analysis. Chicago, IL: University of Chicago Press.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115-135.

Herath, T., & Rao, H. R. (2009). Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. Decision Support Systems, 47(2), 154-165.

Jarvenpaa, S. L., & Majchrzak, A. (2010). Research commentary—Vigilant interaction in knowledge collaboration: Challenges of online user participation under ambivalence. Information Systems Research, 21(4), 773-784.

Jonas, K., Diehl, M., & Brömer, P. (1997). Effects of attitudinal ambivalence on information processing and attitude-intention consistency. Journal of Experimental Social Psychology, 33(2), 190-210.

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. International Journal of e-Collaboration, 11(4), 1-10.

Lapointe, L., & Beaudry, A. (2014). Identifying IT user mindsets: Acceptance, resistance, and ambivalence. In Proceedings of 47th Hawaii International Conference on System Sciences (pp. 4619-4628). IEEE.

Lee, Y., & Larsen, K. R. (2009). Threat or coping appraisal: determinants of SMB executives' decision to adopt anti-malware software. European Journal of Information Systems, 18(2), 177-187.

Li, Y., & Siponen, M. (2011). A call for research on home users' information security behavior. In Proceedings of the 15th Pacific Asia Conference on Information Systems (PACIS). AIS.

Liang, H., & Xue, Y. (2009). Avoidance of information technology threats: A theoretical perspective. MIS Quarterly, 33(1), 71-90.

Liang, H., & Xue, Y. L. (2010). Understanding security behaviors in personal computer usage: A threat avoidance perspective. Journal of the Association for Information Systems, 11(7), 394-413.

Liang, H., Xue, Y., & Pinsonneault, A. (2019). What users do besides problem-focused coping when facing IT security threats: An emotion-focused coping. MIS Quarterly, 43(2), 373-394.

Ludwig, S., De Ruyter, K., Mahr, D., Wetzels, M., Brüggen, E., & De Ruyck, T. (2014). Take their word for it: The symbolic role of linguistic style matches in user communities. MIS Quarterly, 38(4), 1201-1218.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. MIS Quarterly, 35(2), 293-334.

McClain, C., Vogels, E. A., Perrin, A., Sechopoulos, S., & Rainie, L. (2021). The Internet and the pandemic. Washing D.C.: Pew Research Center.

Moody, G. D., Lowry, P. B., & Galletta, D. F. (2017). It's complicated: explaining the relationship between trust, distrust, and ambivalence in online transaction relationships using polynomial regression analysis and response surface analysis. European Journal of Information Systems, 26(4), 379-413.

Moody, G. D., Galletta, D. F., & Lowry, P. B. (2014). When trust and distrust collide online: The engenderment and role of consumer ambivalence in online consumer behavior. Electronic Commerce Research and Applications, 13(4), 266-282.

Moody, G. D., Siponen, M., & Pahnila, S. (2018). Toward a unified model of information security policy compliance. MIS Quarterly, 42(1), 285-311.

Ng, K. C., Zhang, X., Thong, J. Y., & Tam, K. Y. (2021). Protecting Against Threats to Information Security: An Attitudinal Ambivalence Perspective. Journal of Management Information Systems, 38(3), 732-764.

Nunnally, J. C. (1994). Psychometric theory 3E. New York, NY: McGraw-Hill Education.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879-903.

Posey, C., Roberts, T. L., & Lowry, P. B. (2015). The impact of organizational commitment on insiders’ motivation to protect organizational information assets. Journal of Management Information Systems, 32(4), 179-214.

Puhakainen, P., & Siponen, M. (2010). Improving employees' compliance through information systems security training: an action research study. MIS Quarterly, 34(4), 757-778.

Qahri-Saremi, H., & Turel, O. (2020). Ambivalence and coping responses in post-adoptive information systems use. Journal of Management Information Systems, 37(3), 820-848.

Siponen, M., & Vance, A. (2010). Neutralization: New insights into the problem of employee information systems security policy violations. MIS Quarterly, 34(3), 487-502.

Stein, M. K., Newell, S., Wagner, E. L., & Galliers, R. D. (2015). Coping with information technology. Mis Quarterly, 39(2), 367-392.

Straub, D. W., & Welke, R. J. (1998). Coping with systems risk: Security planning models for management decision making. MIS Quarterly, 22(4), 441-469.

DECISION SCIENCES INSTITUTE

Unfolding High-Frequency Trading through the Lens of Responsible AI

Mohotarema Rashid
University of North Texas
Email:mohotaremarashid@my.unt.edu

Md Rasel Al Mamun
University of Old Dominion
Email:m1mamun@odu.edu

ABSTRACT

Algorithmic trading in the financial market has become a norm. High-Frequency Trading (HFT) has drawn attention among scholars due to its unique nature as a part of that algorithmic trading. Slight algorithm tampering can severely impact the financial market's interrelated stakeholders and exchanges, to protect the financial market's stakeholders' interest, comprehensive implementation of responsible AI in the financial market is necessary. As a result, this paper examines the HFT through the lens of responsible AI. Moreover, this paper proposes a framework for responsible AI that includes equity, human control and real-time monitoring, accountability, transparency, and respect for finance fundamentals.

KEYWORDS: High-frequency trading, Financial market, Responsible AI, Algorithmic trading, and Artificial intelligence

INTRODUCTION

Financial Markets that had been previously human-mediated have now become machine-mediated (Angel, 2014). The automated trading system is pervasive in the financial market (Cooper et al., 2020). The ubiquity of technology and the pervasiveness of algorithmic trading enabled a fully automated market. Any trading process that is driven by computers is known as algorithmic trading. In recent years computer-driven trading consists of more than 75% of all trades in the USA (Welcman, 2022; Brogaard, 2010; Hendershott & Riordan, 2013). A subset of algorithmic trading is High-Frequency Trading (HFT). High-Frequency Trading (HFT) is the algorithmic trading technique to complete a sizable number of stock trades in a matter of seconds. More than 70% of the entire dollar trading activity on the U.S. stock market is attributable to high-frequency trading Zhang (2010). Using advanced algorithms, High-Frequency Traders (HFTs) can exploit opportunities emerging from market inefficiencies (Cooper, Davis, & Van Vliet, 2016). Since the effects of the algorithm in the financial market are unpredictable (Wellman & Rajan, 2017), this uncertainty demands more work on a responsible and ethical aspect of Artificial Intelligence (AI), which is the heart of algorithmic trading.

Moreover, technological advancement impacts the questionable role of trading behaviors in the marketplace (Thiagarajan, 2019). As a result, this paper aims to shed light on the HFT through the lens of Responsible AI. To do that, first, we describe why it is essential to look over the HFT from an ethical and responsible aspect; then, we break down the pros and cons of HFT strategies and show how HFT strategies can be used to manipulate the market in addition, we identified the stakeholders and cost bearers of the HFT strategies. Finally, we consider responsible AI's definitions, nature, and aspects and develop a novel framework through which HFT strategies can ensure sustainability in computation and fair and best practices in the algorithmic trading environment.

The stock market is data-driven and heavily based on a complex mathematical algorithm that allows researchers, firms, and data scientists to exploit the data and facilitates decision-making. From this study, researchers, firms, and other associated entities will be able to identify and evaluate software and algorithms in the financial domain, which seem like black boxes to the general people. Since responsible AI in the financial market is an underexplored area in the literature, we hope our study will give new insight into the financial market, computational finance, and data science research.

The interconnectedness of exchange, the use of complex technology, and the pervasiveness of AI pose a threat to the ecosystem of the market. Thiagarajan (2019) illustrated the dark side of reliance on technology and its relationship to HFT. They demonstrated significant incidents that took place in the past and had a substantial impact on the capital market of the USA.
The timelines of some of the significant incidents are given below:

In 2010 the US stock market experienced one of the most devastating incidents in the history of HFT. The leading stock index of the USA, Dow Jones, lost 1000 points within approximately 30 minutes. This reduction in index points is the largest intraday point loss in history. Further, it was explained by the Securities and Exchange Commission (SEC) that the algorithm intended to sell $4.1 billion in future contracts erroneously considered only the contract volume instead of price and time. This type of event compelled scholars, firms, and other associated parties to dig down the mechanism and consequences of algorithmic trading.

Due to an error in the newly installed system, one of the largest market makers in the USA, Knight Capital, lost $460 million in initiating trading on the NYSE. Due to the error in the software, the firm’s computer rapidly bought and sold millions of shares. The company was subsequently forced to take financing and eventually overtaken by its rival.

The Better Alternative Trading System (BATS) is another example of how algorithm errors can adversely impact the market. On issuing its first Initial Public Offering (IPO), the exchange reported that they are having problems related to the auction process. And the error in the system ticker symbol from A to BZZZ was impacted, which caused a decline in Apple’s stock price. After a few moments, the auction problem got resolved, and BATS share went down to $15.25 from its opening auction price, which was $16. Then within 900 milliseconds, BATS stock price reached $0.2848. The exchange withdrew its IPO to stop the drastic fall in the price. The timeframe within which the fall in the price occurred indicates the involvement of HFTs and algorithmic trading.

Due to these alarming events reported above, using the algorithm in trading or using HFT raised debate over the meticulousness and sustainability of those machine learning techniques. The algorithms that are based on AI are the main components of HFT. The incidents reported above made us revisit the role of AI, which is a part of algorithmic trading. Due to the pervasiveness of algorithms and the impact of algorithms in our lives, the "black box" nature of the algorithm poses challenges to tracking its adverse impact on human life (Rahwan et al., 2019). To make the stock exchange more trustworthy to general investors, there should be regulations to monitor the algorithms and software used by institutional investors. Since a positive relationship exists between market volatility and HFT (Ullah et al., 2021), HFT practice can raise tensions among regulators.

To tackle the unwanted behavior of algorithms, we have to rethink the ethical concern of the algorithms. Due to the autonomous agent's opaqueness and unpredictability, HFT may be identified as a threat to the market ecosystem rather than providing social benefits. Researchers call for a new terminology called “Responsible AI” to control and monitor the algorithm in the big data and machine learning era. Accountability, responsibility, and transparency are the three main components of implementing responsible AI (Dignum, 2017). Though automated trading is the norm in the financial market, no single ethical standard still fits all automated trading in the financial market. Davis, Kumiega, and Van Vliet (2013) reported that the regulatory authority of the financial industry, the exchange, has been engaged in creating a fair market. Yet, no single code of ethics is related to external stakeholders in this automated financial era.

To bridge responsible AI from the machine learning field and the financial market, we propose a novel framework that would benefit the financial markets' stakeholders: firms, researchers, trader’s investors, software developers, policymakers, etc. The study aims to develop a novel framework to establish a novel practice in the HFT arena. To shed light on the underexplored area of the implication of responsible AI in HFT, we have explored the following research question: What constitutes responsible AI in HFT practice?

LITERATURE REVIEW

Brogaard (2011) assessed HFT and its impact on the quality of the market. He evaluated the quality of the market in terms of three aspects: price discovery, liquidity, and volatility.
He found that HFT substantially impacts price discovery, provides less liquidity than the other participants, and reduces the market's daily volatility. His key finding was that High-Frequency Traders (HFTs) do not hurt other market stakeholders.

Due to the financial network's interconnectedness and machine trading's domination, Lier (2016) urged for an ethical framework in the financial market. He introduced how morality can be taught to the algorithm and AI. He argued that machines could learn moral behavior by absorbing the information and algorithm that humans give. Even if the system is complex, how a machine learns morals depends on humans.

Mcnamara (2016) combined business, law, and ethics literature and illustrated certain types of HFT practice are unethical, like spoofing and layering (these are discussed in detail later in our paper), and identified four main principles to guide HFT practice; reciprocity, explainability, ensuring a level playing field for all participants in the market and institutional integrity.

Davis et al. (2012) illustrated the ethical implication of algorithmic trading and articulated the necessity of a holistic to see the difference between HFT and regulation. They illustrated that the ethical principles are different for financial analysts (fund managers and investment consultants) and for HFT firms (computer engineers, traders, and quants).

Angel & McCabe (2013) examined fairness in the financial market in Respect of HFT. They illustrated that HFT provides traditional benefits like market making and arbitrage opportunities; in addition, HFT can also be used as a manipulative strategy in the marketplace. Apart from that, they mentioned and evaluated two types of fairness in the market; procedural fairness and fairness in the equality of outcome. Procedural fairness in the marketplace depicts equal treatment of all market participants and fairness in the equality of outcomes such as liquidity and price discovery benefit market participants. As a result, HFT as a whole could not be treated as an unfair practice. They concluded that the definition and use of technology should be differentiated to judge fairness in the marketplace.

Cooper et al. (2016) examined the regulations related to HFT; they articulated that regulations in the financial market are prone to create moral hazard and information asymmetry. As a result, they proposed criteria for HFT to follow. They include employing prudence, which would investigate the risk associated with the strategies; any trading strategies should not hamper or block price discovery.

Ullah et al. (2021) claimed that HFT had become a regular phenomenon in the financial market. They reported around 60% of equities had been traded through HFT in the United States. The authors mentioned that technical innovations could increase productivity and efficiency, but the risk of adopting AI in trading behavior can have drastic consequences compared to the revolution of the technological revolution. They mentioned that since the machine learning method is unambiguous, the outcome can be unexpected. The authors explained responsible AI demands that AI technologies should be used soundly and safely and ensure equity among its users and non-users. They concluded their study by showing a way to achieve socially Responsible investment through implementing responsible AI. The authors added that the existing state of the art is like a black box to the end-users; therefore, an explainable AI is a prerequisite to forming a manipulation-free market.

Mizuta (2022) stated that addressing whether AI can cause market manipulation in the financial sector. Using a genetic algorithm, the author examined whether AI traders discover market manipulations, even if the developer does not intend to manipulate the market. He found that AI traders can discover manipulations that demand attention from financial regulators.

Welcman (2022) points out the ethical concern of HFT in terms of transparency, accountability, and fairness. He mentioned how regulators could act promptly if any trading algorithm makes unexpected changes in its operation. He suggested evaluating the disadvantages of trading algorithms despite their advantages. His future research is directed towards implementing best practices of algorithms precisely when it comes to crisis and uncertainty.

Wellman and Rajan (2017) examined the ethical issues and difficulties raised by the deployment of autonomous trading agents in the financial markets. The writers talk about the possible negative effects of deploying these systems, like how they can affect market stability and efficiency and the moral ramifications of the choices these agents make, like conflicts of interest and justice. To ensure that autonomous trading systems align with societal values and norms, the study emphasizes the significance of developing clear ethical criteria and principles. Though the authors did not mention responsible AI, one possible reason might be that there was no such term in 2017. Still, the measures they have discussed align with the principles of responsible AI.

Han et al. (2022) emphasize using explainable machine learning (XML) methods in high-frequency trading (HFT). The paper makes the case that a new, precise, transparent trading algorithm is necessary, given the complexity and speed of HFT. To increase the readability and transparency of trading decisions made by machine learning models, the author provides a framework for XML in HFT. To provide insights into the dynamics of financial markets and the behavior of market participants, the framework is built on the combination of domain knowledge and machine learning techniques. This research seeks to lessen the likelihood of unethical or undesired trading practices while empowering traders to make knowledgeable, assured judgments.

Scholars in the financial domain and ML domain have been questioning the role of algorithms in the financial market and engaging in research that could lead to ensuring an efficient financial market. My paper is an effort to add insight to the existing literature by shedding light on the employment of responsible AI in sophisticated techniques like HFT in the financial market.

OVERVIEW OF HFT

HFT came into action through an actual legislative change in the United States that was approved by the Regulation National Market System of 2005, also known as Reg NMS (Easley, Lopez de Prado, & O'Hara, 2012). Treleaven (2013) stated that High-frequency trading (HFT) is a distinct area of algorithmic trading where trading activities are carried out in a matter of seconds or even milliseconds. The framework of HFT could not be confined to one domain. Currie and Seddon (2017) defined HFT as a combination of three domains: finance, management, and information system. Since HFT is so successful in terms of profit that Cooper, Davis, and Van Vliet (2016) compared it to gambling.

In the era of algorithmic trading, almost every buy and sell in the financial market takes place through the computer. High-frequency trading is another subset of algorithmic trading (Cooper et al., 2016). Since the financial market is inherently inefficient (Cooper et al., 2016) due to the change that occurs from non-randomness, big firms can exploit that opportunity and ultimately gain from market movement (Cooper et al., 2016). Unlike investors who consider long-term information, HFT traders emphasize short-term information (Cooper et al., 2016).

Speed is one of the major game changers in the HFT technique. As a result, an algorithm that buys and sell with the lowest possible latency is HFT (O'Hara, 2014). O'Hara (2014) came up with an understandable algorithm for HFT. She articulated that Super high-speed computers and co-located servers are the key components of HFT. She illustrated how HFT works; if someone has the information regarding the covariance of all the assets listed in the market, then one can know how security moves through a massive variance-covariance matrix. As a result, if one security goes up, HFT can compute the probability of which security can go up further. Prediction of correlation under different circumstances and strategies is the key algorithm of HFT.

Colocation is another essential feature of HFT. Colocation refers to the traders' server location close to the exchange. Since physical proximity matters in HFT, HFTs invest to put their computer closer to the exchange data center (Angel & McCabe, 2013). Copper et al. (2016) identify computation power, and communication networks are the two key components of HFT. They define HFT as a firm that takes advantage of the price differences between different trading infrastructures.

HFT Provides liquidity in the market (Brogaard, 2010; Easley et al., 2012; O'Hara, 2014), reduces the spread, and excels informational efficiency in the market (Easley et al., 2012). But of the rapid entry and exit nature of HFTs, they are unable to provide constant liquidity in the market (Copper et al., 2016)

Angel and McCabe (2013) reported that HFT provides liquidity and reduces arbitrage opportunities in the market. Still, they cautioned that due to the manipulative strategy, HFT might not offer these benefits (Shorter & Miller, 2014). Since HFT is a part of algorithmic trading, it can reduce volatility, increase liquidity and reduce transaction costs (Davis, Kumiega, & Van Vliet, 2013). By providing liquidity and reducing volatility, HFT can contribute to more price discovery and promote market efficiency. Though HFTs provide liquidity, the liquidity might be so tactical in nature that liquidity can vanish when it is most necessary (Easley et al., 2012). In addition, while interacting among different algorithms, HFT might cause a momentary shortage of liquidity. As a result, the market's volatility might trigger (Angel & McCabe, 2013; Zhang, 2010).

SUPERIORITY OF HFT

Despite being the subset of algorithmic trading, HFT, orders are quickly updated. Due to algorithmic trading, HFTs successfully use the advantage of big data (O'Hara, 2014). Trillions of records are very common in the financial market (Easley et al., 2012). With the help of Machine learning algorithms like K nearest neighbor, multivariate embedded algorithm machines can search for the number of patterns in their recorded events. For example, HFT observes Level III quotes where market participants have in-depth access to the order book. A typical level III order book shows you up to 20 best bids and asks prices with their volume and the associated number of market participants. HFTs understand the exchange matching engine mechanism (O'Hara, 2014). As a result, they optimize their strategy against the exchange matching engine (O'Hara, 2014).

HFTs also employ Natural Language Processing (NLP) tools; they can discover fundamental information by scanning millions of web pages at once (O'Hara, 2014). The market sentiment index that Ravenpack and Reuters develop are the pioneers of using this kind of technology (O'Hara, 2014). By employing machine learning and game theory, HFTs can detect a deep market pattern that is unrelated to any fundamental change in the underline assets (Easley et al., 2012). Complex algorithms associated with sophisticated tools put HFTs ahead of other market participants. Currie and Seddon (2017) illustrated HFTs don't use ordinary database query tools; it was seen in HFT conferences where participants were presented using languages such as C, C++, Perl, and Fin QL (Financial version of SQL).

To make money in the algorithmic market, technology speed is a prerequisite, and HFTs ensure that speed which enables them to make a profit and avoid loss (Cooper et al., 2016). Colocation and exchange matching strategy enable HFT traders quickly respond to any substantial information change.

To sum up we can say complex algorithms, super speed, colocation with the exchange, and computing power are the fundamental components of HFT, which make HFT superior to other trading strategies.

MARKET MANIPULATION THROUGH HFT

Two major components of an efficient market are liquidity and less volatility. Scholars have differed in their view on whether HFT consistently provides liquidity and volatility. The algorithm of HFT can sometimes trigger microstructure mechanisms with probable foreseeable results (O'Hara, 2014). Below we are illustrating some of the market manipulative techniques that can be done through HFT:

Front Running

Known as a predatory algorithm that can identify that a large amount of share order is applied for and attempt to move in front of it; as a result, HFT gains at the expense of others (Angel & McCabe, 2013).

Quote Stuffers

The sole intention is to slow down the exchange of unnecessary messages (Angel & McCabe, 2013; Easley et al., 2012; O'Hara, 2014).

Quote dangling

Quote danglers enter and cancel limit orders instantly with the goal of obscuring the quoting process (Easley et al., 2012; O'Hara, 2014; Shorter & Miller, 2014).

Pack Hunters

Several HFTs intentionally know and track each other's activity and exploit the opportunity if there is any (O'Hara, 2014).

Liquidity Squeezers

When a large number of investors try to reverse their position, they trade in the same direction and take liquidity as much as they can. As a result, price spikes and profit is realized (Carlin, Lobo, & Viswanathan, 2007; Easley et al., 2012; O'Hara, 2014).

Algo-Sniffing

In this process, HFT can detect the presence of competing algorithms and execute trades faster than the competitors (Boatright, 2013; Davis et al., 2013).

Order Triggering Strategies

Traders intentionally place large short sale order to depress the price and then buys the stock back when the price is really low, through which they cover the short sale and realize profit; the HFTs can take advantage of this type of situation (Angel & McCabe, 2013).

Spoofing

Through this, a trader who intends to buy might place a sale order to deceive other investors into trading (Angel & McCabe, 2013). For example: if a large number of shares is available at the offer price and an investor wants to buy 4,000 shares with a bid price of $20 and an asking price of $20.02. An investor who knows trading decisions is taking place in full automation set a sell order at $20.01 with a view to convincing the algorithm to reduce the offer price. When the offer price falls to $20.01, the investor cancels his own sell order and quickly places a large buy order at the lower price. As a result, the investor can save $0.01 on the buying price.

Wash Sales

A false sale is framed to give an impression of higher trading volume in a stock market than the actual volume (Angel & McCabe, 2013).

STAKEHOLDERS AND COST BEARERS OF THE HFT

Below we illustrated some of the key stakeholders of HFT:

LFTs

LFTS stands for Low-Frequency Traders, opposite to the High-Frequency Traders. Like HFTs, LFTs also provide liquidity and act as market makers. They provide market-making services by submitting buy and sell limit orders (Cooper et al., 2016). Generally, they are appointed by the exchange. Trading against investor orders allows them to earn from the spread, which is the price difference between the bid and offer prices. Market makers have historically been the main source of liquidity for investors, adjusting the amount of the spread based on their assessment of risk and competition from other market participants (Davis et al., 2013).

Investors

Investors are the people who are interested in fundamental information change in securities, and their interest lie on the long-term horizon. Their sole concern is making a profit by buying and selling securities in a fair environment (Cooper et al., 2016).

Computer engineers

They consist of software engineers and computer scientists. Their main aim is to optimization of technology. They support infrastructure, business logic, and algorithms developed by traders and quantitative analysts and are prioritized for rapid implementation by developers and network engineers (Davis et al., 2013; Lehmann, 2013).

Quants

They focus on mathematical constructs and data. The priority is given to mathematical precision, adherence to laws and trading rules that act as boundaries, and ethical issues are sometimes eclipsed in this discipline (Cooper et al., 2017; Davis et al., 2013).

Apart from the above significant stakeholders, regulators and society as a whole are the stakeholders of HFT (Lehmann, 2013). Generally, Low-Frequency Traders are the cost-bearers of HFT. O'Hara (2014) reported that if HFTs can figure out what LFTs are doing, they can take advantage of that. For example, if an LFT places an order up and sends it in initiating of each minute, a larger percentage of trades will be executed at the first moment of the minute, and HFTs can easily discover this strategy.

LFTs are bound to do market research which is fundamental to price discovery in the market, but due to the advanced technology, HFTs do not have to do this type of research. Front running refers to trading for yourself instead of customers, which is considered illegal in the USA; since HFT doesn't work for anyone, they can easily engage in front running without having any liability (Cooper et al., 2016). LFTs' liquidity and profitability are also squeezed by HFTs (Cooper et al., 2017).

Angel and McCabe (2013) pointed out that if the connected network of the HFTs goes slightly wrong, then the consequences might be like the Flash crash of May 6, 2010. As a result, HFTs pose an excessive risk to market participants. The uncertainty in the algorithm can trigger events like the flash crash of 2010, the exact cause of which seems unknown till a couple of years. And then, it was revealed that the crash occurred due to a manipulative strategy like spoofing. Kirilenko and Lo (2013) investigated unexplained events in financial markets, including flash crashes, and concluded that due to the pervasiveness of HFT, prices could change rapidly, which could create more tension among market participants.

RESPONSIBLE AI

Since the intelligent system has been making decisions in this big data and machine learning era, this decision impacts human life; it is high time to rethink the ethical implication of AI (Dignum, 2017). The promise of practical, affordable, or "neutral" big data solutions has led public organizations to employ algorithmic systems to deliver public services (Busuioc, 2021). But the algorithm is a chain of rules to achieve the desired outcome. The training set algorithm can learn complex features and predict future events. How accurately it can predict the future that depends on the training set. In the case of deep learning algorithms, how they reach the ultimate output is often not transparent (Adabi & Berrada, 2018; Busuioc, 2021). As a result, the deep learning algorithm is often compared to the black box algorithm. Guidotti et al. (2018) said that the black box algorithm of deep learning might seem opaque but more powerful than simple ML models (Guidotti et al., 2018). They added that companies often release services and solutions in safety-critical industries using machine learning and data mining. However, the training data's spurious correlations or artifacts pose a risk of systematic bias. Technologies for explanation can help businesses make products that are safer and more reliable, as well as handle any potential liabilities. Below I have illustrated some major components of responsible AI from different fields:

Table 1: Components of Responsible AI from literature

Authors	Components of Responsible AI
Guidotti et al. (2018)	Interpretability, Accuracy, Fidelity
Dignum (2017)	Transparency, Accountability, Responsibility
Busuioc (2021)	Transparency, Accountability, Explanation, Justification
Adabi and Berrada (2018)	Systematic Definition, Abstraction, quantification of Explicability
Floridi et al. (2018)	Beneficence, Non-maleficence, Justice, Explicability
Cowls, King, Taddeo, and Floridi (2020)	Falsifiability and gradual implementation, prevent predictor manipulation, Interventions that take the receiver's context into account, transparent, protecting privacy, situational fairness, human-friendly semantics
Clarke (2019)	Evaluate positive and negative impacts with implication, Complement and ensure human control, Healthy and safe human protection, respect for human rights and values, ensure accountability and transparency, Combine Quality Control, Demonstrate Strength and Adaptability, Ensure Responsibility for Duties Impose and Accept Penalties and Sanctions
Asilomar AI Principle (2017)	Safety, Transparency, Responsibility, Value alignment, Human Control, Non-subversion
Buruk, Ekmekci, and Arda (2020)	Human Control, Autonomy, Transparency, Security, Utility, equality
Peters, Vold, Robinson, and Calvo (2020)	Research, Insight, Ideation, Prototype, Evaluation
Zeng, Lu, and Huangfu (2018)	Humanity, Collaboration, Share, Fairness, Transparency, Privacy, Security, Safety, Accountability,

Table 1: Components of Responsible AI

Guidotti et al. (2018) attempted to shed light on the explanation of the black box system of the algorithm and urged for three components to decipher the black box; interpretability, accuracy, and fidelity.

Dignum (2017) articulated that responsible AI is more than ticking ethical values in a tick box and developed the famous ART model of responsible AI. ART model includes transparency, accountability, and responsibility of the autonomous system. He emphasized that the owner and users of the AI system are the full controllers of the AI system, and the control should prevail. Busuioc (2021) developed an accountability framework through which he stated components through which responsible algorithms can be implemented; his framework includes; transparency, accountability, explainable, and justifiable.
Adabi and Berrada (2018), through a survey of 381 papers and proposed formalized approach to ensuring explainable AI. Their approach includes systematically defining common terms used in different domains, developing a generic explainable framework for stakeholders, and incorporating formalized evaluation metrics.
Floridi et al. (2018) propose a new component to traditional bioethics principles: Beneficence, Non-maleficence, Autonomy, Justice, and Explicability to form a good AI society collaboration with twelve AI experts.
Cowls et al. (2020) again addressed seven essential factors for implementing AI for social good. Clarke (2019) illustrated the principles of responsible AI by elaborating on 10 (Table:1) themes that are derived from 50 principles. His aim was to suggest organizations that want to employ responsible AI within the organization and address the welfare of stakeholders. He identified five responsible entities in every five phases of implementing AI; research, creation, improvement, sharing, and implementation. The distinct thing about his theme is that it implicitly considers the interest of users, owners, business firms, and its associated stakeholders.
In 2017, experts in AI that, include researchers, industrialists, and scientists, gathered at a conference arranged by future life institute and initiated 23 principles known as Asilomar AI principles (Asilomar, 2017). Asilomar AI principles illustrated 23 principles for implementing beneficial AI (Asilomar, 2017). That principle has three aspects; research, ethics and values, and longer-term issues related to AI.
Buruk et al. (2020) examined three documents that comprehend the moral aspect of AI: The Montréal Declaration for Responsible Development of Artificial Intelligence, Ethics Guidelines for Trustworthy AI, and Asilomar Artificial Intelligence Principles. They found five broad themes by analyzing these three documents; human control, autonomy, transparency, security, utility, and equality.
The P7000 standards project was started by the IEEE in 2019 to address ethical issues surrounding the development of autonomous and intelligent systems. Peters et al. (2020) illustrated that the principles that IEEE has initiated might face challenges while turning into practices. As a result, they came up with a responsible system design process by incorporating five explainable steps (Table 1).
Zen et al. (2018) examined AI principles from academia, industry, and government and developed keywords that depict the prior mentioned entities that adopted AI principles. Both academic computer science and professional software and computer engineering place emphasis on software quality and fail-safe design, and ethics are interwoven into the rules of software and hardware design and testing (Cooper et al., 2020).
The prior scholarly work on responsible AI suggests an integrated system where transparency, fairness, equitability, and accountability are the core components. In addition, respect for people’s privacy, security, and interests are the frequently discussed components of responsible AI. There is no one set of principles applicable to all domains; the impact of each element on the direction of responsible AI can vary across disciplines.

HFT THROUGH THE LENS OF RESPONSIBLE AI

The idea is that algorithms function without human interpretation and are objective and accurate. The algorithm removes emotion from decisions (Currie & Seddon, 2017). Due to its ability to remove subjectivity and bias, this leads to the hypothesis that AI decision-making is more trustworthy than human decision-making alone, especially when buying or selling stocks (Martin, 2019). But the flash crash 2010 compelled us to question the fairness of employing machine learning techniques in the financial world (Welcman, 2022). Incidents like the Flash Crash 2010 opened our eyes by identifying the exact reason for the crash, which depicts the opaqueness and uncertainty of the HFT algorithm (Currie & Seddon, 2017). Cooper et al. (2016) illustrated that smarter and faster competitors consistently outperform shallow and slow competitors. They explained some adverse possibilities, such as losing control over the HFT strategy, not taking direct responsibility for providing liquidity, and a secret profit-making method that can create stress in the financial market. If HFTs are associated with a market-making strategy without harming the price, this feature does not violate the principles of responsible AI.

Below I have discussed several issues through which we can answer whether HFT is a barrier to implementing responsible AI in the financial market:

Advantages of computing power

Speed, computation power, and complex algorithm are three game-changing components of HFT. Since quants and computer engineers are experts in this trading area, they modify their algorithms to match their strategy in an exchange matching engine (Easley et al., 2012). Moreover, Since HFTs can detect human activities in the system, they are identified as predator algorithms (Easley et al., 2012; O'Hara, 2014). HFT pays a premium to have colocation with the exchange to ensure profit. As a result, lack of computation power and money might affect the fair practice in exchange.

Technology and information asymmetry leads to inequality

Technology and information asymmetry are two main concerns of HFTS (Currie & Seddon, 2017). One of the fundamental principles of responsible AI is equity, but LFTs and market intermediaries claim their profit is being taken by HFTs (Cooper et al., 2017). As a result, Cooper and Colleagues (2016) urged for a platform (market) where reliability and interpretability will be ensured. They also illustrated the arms race among business firms that employ sophisticated technology to survive in the competition. Information asymmetry is an inherent characteristic of the financial market; due to the adoption of sophisticated technology by HFTs, technology asymmetry, and information asymmetry have become the norm. That violates the equity principle (Currie & Seddon, 2017). Zook and Grote (2017) identified information inequality that causes an unfair market environment. Since speed and algorithm superiority drive the market, HFTs can gain with the cost of LFTs (Davis et al., 2013).

Davis et al. (2013) mentioned quality arbitrage if an entity has better speed, better technology, and better algorithmic strategy, then they can make a profit by employing that better quality over other traders. As a result, in terms of the selection and execution of orders, HFTs can gain more profit. Since HFT is competitive, firms don't even discuss their program languages (Currie & Seddon, 2017).

Due to the nature of technology and information asymmetry inequality might prevail in the financial market. Moreover, colocation is the source of inequality which is one of the vital components of HFT (Angel & McCabe, 2013; Welcman, 2022). Since not every investor has the ability to invest in powerful computers and technology, high-speed players are successful in getting information compared to slower ones (Mackenzie et al., 2012; Welcman, 2022).

Protecting Stakeholder's Interest

Currie and Seddon (2017) reported that Not all HFTs are harmful, but HFTs who involve in manipulative strategies instigate unfair practices by using algorithms. They illustrated real-time monitoring, expectations over their trade, and reversing position if a loss occurs; all these components should be employed to employ responsibility in the market. They argue that the market should be a safe place for long-term investors. Since the financial market has several stakeholders, the principles of AI should be considered; for example, Nyholm and Smids (2016) illustrated that while designing a self-driven car, programmers should keep in mind the pedestrians, traffic, humans, and so forth. Currie and Seddon (2017) reported irrespective of HFTs' dominance in the market, they are the heart of the market, and long-term investors' interests have not been taken care of. They urge for more collaboration between finance, management, and information system to unfold the practice of HFT. Cooper et al. (2020) illustrated four qualities of an efficient market: voluntary participation of market participants, transparency regarding quotes and volume of the market, informational efficiency, that is, the price reflects all available information and reliability, which is a low possibility of loss. Cooper et al. (2016) reported that HFTs enjoy more privilege than LFTs; HFTs have direct access and withdrawal option regarding their position, whereas LFTs does not have this sort of privilege. One market participant should not interfere and take away other market participants' profit (Cooper, Ong, & Van Vliet, 2015).

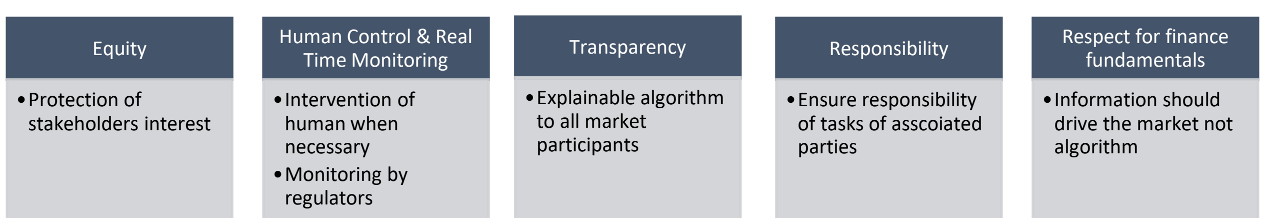
Explicability of the HFT Algorithm

HFT strategy should be prudent; prudent means the strategy risks are fully understood and investigated (Cooper et al., 2015). Regulators should not allow trading methods to hide the algorithm affecting market transparency. Sometimes, the algorithm employed by HFTs is so complex that Currie and Seddon (2017), through an empirical study, showed that some technologists and even traders do acknowledge that regulators are not as experts as HFTs and quants.

A FRAMEWORK FOR EMPLOYING RESPONSIBLE AI IN HFT

Heilinger (2022) argued technical issues and ethical issues should be addressed together. By aligning previously discussed components of responsible AI, we have proposed a model of a responsible framework for HFT

Figure 1: Framework of responsible AI in HFT



Equity

Since HFTs constitute a significant part of the financial market (Sobolev, 2020), equity among HFTs and Non-HFTs should be ensured. HFT is one of the most complicated parts of algorithmic trading; in the case of other market participants, before understanding what is going on in the market HFTs realize their profits; as a result, information and technological asymmetry increase. For the above reason, HFT practice might violate one of the key principles of responsible AI: ensuring equity between users and non-users of technology. Computation power and colocation strategies put HFTs way ahead of general investors.

Human Control and real-time monitoring

Though algorithms can remove human emotions from decisions as a part of that HFTs response to the time of market inefficiency, AI should enable human intervention when necessary. Past algorithmic errors in exchange made it imperative that the algorithm be simple or complicated, and AI still can not replace human judgment in critical decisions. Cooper et al. (2015) suggested three requirements for algorithmic trading: statistical control, acceptable risk and loss behavior, and real-time monitoring to control unexpected behavior. The algorithm should have human control and enable real-time monitoring to ensure safety in terms of crisis.

Transparency

HFT is famous for its opaqueness in algorithms and complexity in order systems. As a result, HFT might violate the transparency principle of responsible AI. The Explicability of the algorithm ensures fair practice in machine learning. Super speed and complex algorithms pose challenges for other key market participants and regulators to track the activities of HFTs. As a result, fair practice in the market might get hampered, and one market participant might gain at the expense of another market participant.

Accountability

Since there are some key participants in the market, brokers, traders, and money managers, they are bound to follow the decorum of the stock exchange (Davis et al., 2013). To ensure the accountability of each participant in the market, enablers of automated agents like NASDAQ and NYSE should come forward (Davis et al., 2013). We argue that software developers and computer engineers should have accountability regarding algorithms. Apart from them, the exchange itself and the regulatory authority of financial industries have been implementing ethical standards in the marketplace so that general people can rely on the market (Davis et al., 2013).

Respect for fundamentals of finance

We should not forget the fundamentals of finance in the presence of advanced technology. One must defend the argument that we neglect financial experts and prefer quants and computer scientists over them. Though mathematicians, computer scientists, and software engineers employ market efficiency, they tend to have little finance knowledge (Cooper et al., 2016). The Efficient Market Hypothesis (EMH), put forth by Fama in 1970, contends that the stock price should accurately reflect all information available to the market. Fundamental information change should be the key driver of the financial market since security prices should only reflect all available information. As a result, finance fundamentals should be given importance along with speed and computation.

Easley et al. (2012) contrasted a significant difference between High-Frequency Traders (HFTs) and Low-Frequency Traders (LFTs). They showed a typical financial analysis conference participants are LFTs whose subject matters of discussion concentrated on fundamentals analysis like monetary policy, asset allocations, and stock and financial statement valuations. In contrast, HFT conference participants are computer scientists who discuss machine learning algorithms, TCP/IP connections, lowest latency, colocation techniques, and, most importantly, exchange matching engines.

CONTRIBUTIONS

In this paper, we shed light on the HFT concerning responsible AI. Reviewing different components of responsible AI, we propose a model through which the HFT practice can behave responsibly. The features of the model include equity, human control and real-time monitoring, transparency, responsibility, and respect for the fundamentals of finance.

The notion is not straightforward whether HFT practice is a threat or a blessing to the financial ecosystem. We have seen those certain types of HFT practices can catalyze market manipulation. However, HFT provides more liquidity, price discovery, and reduced volatility. On the journey of demystifying HFT through the lens of responsible AI, we have realized that in the race for speed and computation power, if small players like (LFTs) are wiped away, the financial ecosystem would get harmed. Thus, the framework will help practitioners to decide on using HTF in the financial environment as well as will expand the body of knowledge to the HTF concerning AI.

CONCLUSION

AI facilitates human decision-making in the market ecosystem by removing time and place barriers. But some prominent claims against HFT, like gaining at the expense of others and opaqueness in the algorithms, compelled us to think about equity and fairness in the marketplace. Although Welcman (2022) said that since programmers develop algorithms, they are solely responsible for ensuring ethical standards, still to ensure a responsible environment throughout the financial market, all key market participants should be brought under one umbrella of a master framework described above. Since in the financial market, there are several stakeholders and exchanges are interconnected with each other, a slight manipulation of an algorithm can create a significant impact on the market ecosystem. Employing responsible AI in the regulation of autonomous agents in the financial domain is crucial and may contribute to the solution for addressing the problem of control of AI.

REFERENCES

Adabi, A., & Berrada, M. (2018). Peeking inside the Black-Box: A Survey on Explainable Artificial Intelligence. *IEEE Access*, 6, 52138-52160.

Angel, J. J. (2014). When finance meets physics: The impact of the speed of light on financial markets and their regulation. *Financial Review*, 49(2), 271-281.

Angel, J. J., & McCabe, D. (2013). Fairness in financial markets: The case of high frequency trading. *Journal of Business Ethics*, 112, 585-595.

Asilomar Conference on Beneficial AI. (2017). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Asilomar_Conference_on_Beneficial_AI&oldid=1095395515

Boatright, J. R. (2013). Ethics in finance: *John Wiley & Sons.*

Brogaard, J. (2010). High frequency trading and its impact on market quality. Northwestern University Kellogg School of Management Working Paper, 66.

Buruk, B., Ekmekci, P. E., & Arda, B. (2020). A critical perspective on guidelines for responsible and trustworthy artificial intelligence. *Medicine, Health Care and Philosophy*, 23(3), 387-399.

Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825-836.

Carlin, B. I., Lobo, M. S., & Viswanathan, S. (2007). Episodic liquidity crises: Cooperative and predatory trading. *The Journal of Finance,* 62(5), 2235-2274.

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, 35(4), 410-422.

Cooper, R., Davis, M., Kumiega, A., & Van Vliet, B. (2020). Ethics for automated financial markets. *Handbook on ethics in finance*, 1-18.

Cooper, R., Davis, M., & Van Vliet, B. (2016). The mysterious ethics of high-frequency trading. *Business Ethics Quarterly*, 26(1), 1-22.

Cooper, R., Ong, M., & Van Vliet, B. (2015). Multi-scale capability: A better approach to performance measurement for algorithmic trading. *Algorithmic Finance*, 4(1-2), 53-68.

Cowls, J., King, T., Taddeo, M., & Floridi, L. (2020). How to design AI for social good: seven essential factors. *Sci Eng Ethics*, 26, 1771-1796.

Currie, W. L., & Seddon, J. J. (2017). The regulatory, technology and market ‘dark arts trilogy’of high frequency trading: a research agenda. *Journal of information technology*, 32(2), 111-126.

Davis, M., Kumiega, A., & Van Vliet, B. (2013). Ethics, finance, and automation: A preliminary survey of problems in high frequency trading. *Science and engineering ethics*, 19, 851-874.

Dignum, V. (2017). Responsible artificial intelligence: designing AI for human values. *World*, 13, 23.

Easley, D., Lopez de Prado, M., & O'Hara, M. (2012). The volume clock: Insights into the high frequency paradigm. *The Journal of Portfolio Management,*(Fall, 2012), Johnson School Research Paper Series(9-2012).

Fama, E. F. (1970). Efficient market hypothesis: A review of theory and empirical work. *Journal of Finance*, 25(2), 28–30.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . Rossi, F. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, 28, 689-707.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42.

Han, H., Forrest, J. Y. L., Wang, J., Yuan, S., Fei, H., & Li, D. (2022). Explainable Machine Learning for High-Frequency Trading Dynamics Discovery (*SSRN Scholarly Paper No.* 4256777). https://doi.org/10.2139/ssrn.4256777

Heilinger, J.-C. (2022). The ethics of AI ethics. A constructive critique. P*hilosophy & Technology*, 35(3), 61.

Hendershott, T., & Riordan, R. (2013). Algorithmic Trading and the Market for Liquidity. *Journal of Financial and Quantitative Analysis*, 48(4), 1001–1024.

Kirilenko, A. A., & Lo, A. W. (2013). Moore's law versus murphy's law: Algorithmic trading and its discontents. *Journal of Economic Perspectives*, 27(2), 51-72.

Lier, B. van. (2016). From High Frequency Trading to Self-Organizing Moral Machines. *International Journal of Technoethics*, 7(1). https://doi.org/10.4018/IJT.2016010103

Lehmann, A. G. (2013). Quality and governance in high frequency trading systems. [Master's thesis, University of Oslo].

Mizuta, T. (2022). Artificial intelligence (ai) for financial markets: a good AI for designing better financial markets and a bad ai for manipulating markets. In Digital Designs for Money, Markets, and Social Dilemmas (pp. 305-329): *Springer.*

MacKenzie, D., Beunza, D., Millo, Y., & Pardo-Guerra, J. P. (2012). Drilling Through the Allegheny Mountains. *Journal of Cultural Economy*, 5(3), 279–296. https://doi.org/10.1080/17530350.2012.674963

McNamara, S. (2016). The Law and Ethics of High-Frequency Trading. *Minnesota Journal of Law, Science and Technology,* 17(1), 71–152.

Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? *Ethical theory and moral practice*, 19(5), 1275-1289.

O'Hara, M. (2014). High-frequency trading and its impact on markets. Financial Analysts Journal, 70(3), 18-27.

Peters, D., Vold, K., Robinson, D., & Calvo, R. A. (2020). Responsible AI—two frameworks for ethical design practice. *IEEE Transactions on Technology and Society*, *1*(1), 34-47.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J., Christakis, N., Couzin, I., Jackson, M., Jennings, N., Kamar, E., Kloumann, I., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D., Pentland, A., & Wellman, M. (2019). Machine behaviour. *Nature*, 568, 477–486. https://doi.org/10.1038/s41586-019-1138-y

Sobolev, D. (2020). Insider Information: The Ethicality of the High Frequency Trading Industry. *British Journal of Management*, 31(1), 101–122. https://doi.org/10.1111/1467-8551.12366

Shorter, G. W., & Miller, R. S. (2014). High-frequency trading: background, concerns, and regulatory developments (Vol. 29): Congressional Research Service Washington, DC, USA.

Thiagarajan, R., Lacaille, R. F., Im, H., & Wang, J. (2019). The Need for Speed: Does High-Frequency Trading Make or Break Equity Markets? *The Journal of Index Investing*, 10(2), 6–23. https://doi.org/10.3905/jii.2019.10.2.006

Treleaven, P. G. (2013). Algorithmic Trading Review. *Communications of the ACM*, 56(11).

Ullah, A. K. M. A., Sultana, S., Faisal, F., Rahi, M., Alam, Md. A., & Alam, M. G. R. (2021). A Brief Review of Responsible AI and Socially Responsible Investment in Financial and Stock Trading. https://doi.org/10.36227/techrxiv.15093762

Welcman, Y. (2022). Ethical Use of Algorithms in Times of Uncertainty and Crisis: The Case of High-Frequency Trading. *Available at SSRN* 4179166.

Wellman, M. P., & Rajan, U. (2017). Ethical issues for autonomous trading agents. Minds and Machines, 27, 609-624.

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking artificial intelligence principles. arXiv preprint arXiv:1812.04814.

Zhang, F. (2010). High-Frequency Trading, Stock Volatility, and Price Discovery*. SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.1691679

Zook, M., & Grote, M. H. (2017). The microgeographies of global finance: High-frequency trading and the construction of information inequality. *Environment and Planning A: Economy and Space*, 49(1), 121-140.

Decision Sciences Institute

Use of Process Mining Analytics to Overcome Deficiencies of Process Modeling Practice

Julian Koech
Illinois State University
Email: jckoec1@ilstu.edu

Borinara Park
Illinois State University
Email: bpark@ilstu.edu

ABSTRACT

This paper presents a way of combining process mining analytics with process modeling in order to improve the quality of input data for the process simulation. Process modelers often use insights from traditional sources such as process documentations, observations, and interviews with stakeholders. Obtained information from these sources likely presents a biased and inaccurate view of the real process. Instead, in this paper, it is demonstrated that the process-related components such as activities, relationships, durations, and resources are extracted from process mining and used as input data for the creation of a process model that closely reflects reality. The proposed combination is applied in a case study of a purchasing requisition process.

KEYWORDS: Process mining; Process modeling; Discrete Event Simulation; Process Mapping

INTRODUCTION

Process modeling is a technique used in the analysis of complex business processes. Such business processes contain a series of connected entities that use the organization's resources to produce a specified output for a particular customer or market (Doomun & Vunka Jungum, 2008). Process modeling technique is therefore used to combine and define the relationship between these critical entities and resources, while its analysis produces key performance data outcomes that can be used to understand their cause and effects relationships. It allows managers to understand the fundamentals of business processes, identify opportunities for change and assess the impact of proposed changes (Aguirre et al., 2013). It is mostly used in industries like manufacturing, healthcare, transportation, and defense, to optimize processes, improve efficiency, and decision making (Solding, P., & Gullander, P. 2019). As such, it has become an essential tool for businesses and organizations looking to better their operations and gain a competitive advantage in their respective industries. Process modeling technique, however, relies heavily on the type of information used to create it. Making it susceptible to the "garbage in garbage out" phenomenon.

Most often, process modelers create models based on insights from documenting processes, interviews with stakeholders, and directly observing a process (Rozinat et al., 2009). Obtained modeling information through these sources might present a biased and inaccurate view of the actual process, resulting in disparity between the actual behavior of the process and the developed model (Abohamad et al.2017). Lynch et al. (2014) supports this by stating that modelers are often challenged to accurately model a system that closely reflect the reality. However, advancements in technology have enabled business processes to be supported by information systems such as Enterprise Resource Planning (ERP) and Supply Chain

Management (SCM) which extract organizational data in the form of event logs that can be analyzed in process mining packages (Aguirre et al., 2013). Process mining is a recent discipline that generates useful information that organizations can use to understand and improve their business processes (van der Aalst, W.M.P., et al., 2011). It involves analyzing event log data to gain insight to patterns and trends in a process and understand how this impacts current performance (Van Der Aalst et al.,2004). Such knowledge from event logs can support modeling activities in the conceptual process modeling phase (Abohamad et al. 2017).

Research from Aguirre et al., (2013), Van Der Aalst, (2011), Weske, (2012) and Jadrić et al., (2020) has shown that the integration of process modeling with other techniques such as process mining would help solve process modeling's challenge on input data needed to model accurate processes. Despite the growing number of studies relating to the coupling of process modeling with process mining, challenges persist in the research area which needs to be addressed (Van Der Aalst, 2012, Zakarija et al., 2020). For instance, the focus of most studies lacks clear guidelines on how process mining data can be extracted to support the construction of process models. The methods used to extract event log data are general and abstract, without results from real case study (Wang et al., 2018).

This paper presents the development of a process model using the process mining information to support the creation of the model that is less biased and has a close representation of the real process. The focus is to illustrate how process mining information can be extracted and used as input for process model development. The primary aim is to overcome the limitation of the process modeling practice where subjective information from the modeler is the main source of model inputs. A case study of a purchasing requisition is used to demonstrate how the actual data about the process has been extracted and transformed to create a process model with objective inputs.

LITERATURE REVIEW

The literature review focuses on process modeling previous practices, and combination efforts of process modeling and process mining.

Process Modeling Practices

A process model is an abstract description of an actual or proposed process. It represents selected process elements considered important in the model's creation. Abohamad et al. (2017) stated that, to develop a validated process model, input data used in the creation of the model is a basic entity. Robinson et al (2012) created a Discrete-event simulation (DES) model of a hospital process that modeled the flow of patients through an outpatient theater over a day. They obtained information from interviewing senior executive managers, senior medical consultants, departmental managers, nurses, support workers, ward clerks and staff.

This data was used to validate the integration of Discrete-event simulation (DES) with lean methods. An approach they called SimLean. They modeled the hospital process which generated results on the use of resources, size of queues and the number of patients in the system (including arrivals and discharges). The model was based on the SimLean model in which its success was measured through a meeting with the lean team. They mentioned that their aim was not to develop an exact representation of the real process, but to set process names and the model data (quantities of resources and timings) to approximately the right level.

Mahmood et al. (2019) used simulation process modeling software, AnyLogic to build and simulate dynamic models for engineering systems. Input data for the process model was collected from a software called Simuland and from observations of how the process worked. Abohamad et al. (2017), objects this method of data collection by stating that modeling processes based on insights from observation might present biased and inaccurate view of the process.

Doomun & Vunka (2008) also developed a process model of a call center. They used input data recorded by computers called Automatic Call Distributor (ACD) and Interactive Voice Response (IVR) which used data to attribute the process flow of calls. The computers recorded the call's identification number, action taken, time elapsed since the earlier action, numbers of arrivals and abandonment, average service times, agent utilization, and the distribution of delays in a queue. They used quantitative analysis of the data obtained to measure the performance of the system. They faced a challenge where they lacked modeling data on the relationship between call-by-call stored at the IVR level. They recommended that data mining techniques can be applied to help mitigate this problem.

In support of this, process mining founder, der et al (2013) confirmed that event log data from process mining can support process modeling activities. He explained that event logs from information systems can be analyzed using process mining techniques, which help handle discrepancy between how processes operate in real life and process guidelines. Abohamad et al. (2017) adds that process knowledge discovered from analysis of event data can support process modeling activities. The process knowledge can then be cross-checked against traditional sources of information.

Attempts to combine Process Modeling and Process Mining

Abohamad et al. (2017) presented a hybrid framework within a hospital's emergency department that integrated process mining to support process modeling activities as part of the steps in developing simulation models. They used historical data, in an event log structure, provided by managers over one year. Using ProM and Disco software, they obtained statistical data in the conceptual modeling phase. They then used AnyLogic software to develop a discrete event simulation model, from which they found satisfactory results in comparison to the key performance indicators of the historical data. He recommended that efforts are needed to capture interaction between event log cases and extraction of resource schedules, queues, and activity duration for process modeling.

Măruşter & van Beest, (2009) proposed a process redesign methodology based on coupling process mining and process modeling. From three case studies, they used Colored Petri Net (CPN) simulation to simulate resource usage optimization and throughput time. The focus on this study, however, was on CPN simulation, which emphasized understanding of the first stage of process redesign project. Aguirre et al., (2013) complemented this work by putting emphasis on project understanding phase. They used a procurement process as a case study in which event log data was extracted from Enterprise Resource Planning (ERP) system. The information supplied by the log included case id, time stamps, activities and performers. Through the application of process mining algorithms such as alpha mining (van der Aalst et al., 2004), it was possible for them to automatically discover the actual process model using the ProM software functionality, and later represented it in a Petri Net.

They then simulated and evaluated different process improvement options based on waiting times calculated through the statistical analysis of the event log data. In their findings, the ProM petri nets used were not easy to interpret, especially for business users. They recommended other process mining tools such as Disco from Fluxicon that would supply more understandable visualization and animation of event log data. A general approach was proposed by Jadrić et al. (2020) where they focused on possible issues and advantages of the detection of student behavior and processes based on the data from a standard learning management system. He proposed the potential of using process mining results as input data for discrete-event simulation modeling in an educational sector.

PROBLEM STATEMENT

Despite preliminary research efforts to integrate process mining and process modeling, there is lack of clear guidelines as to how process mining data can be extracted and used as input data for constructing process models. Martin et al. (2015) attempts to bridge this gap by suggesting a structured overview of process modeling and how process mining can support the modeling tasks. They do not, however, show how these guidelines can be practically applied. Abohamad et al. (2017) even adds that most reported research neglects the internal structure of process modeling and how it is related to the output of process mining algorithms. This paper's main goal is to explore and demonstrate specific ways to improve the process modeling practice by integrating it with process mining information, ultimately resulting in better process models that better reflect the reality.

METHODOLOGY

A purchasing requisition process is used as a case study for the proposed framework. The event log of the actual execution of the purchasing requisition is obtained in an excel file. It is then imported to the process mining software, Disco, for analysis. This software is preferred due to the complexity of the proposed case study (having 9,120 cases) and its unstructured nature. Disco produces fuzzy miner algorithms or process maps which are suitable for analyzing such complex processes at various levels of granularity. It also helps provide meaningful abstraction and different views of the process (Günther & van der Aalst, 2007).

For a successful import of the log data in the process mining tool for analysis, the data is required to have at least a Case ID, Activity name, Timestamp and Resources (Rozinat et al., 2009). The Case ID associates an activity with a particular case, for example, client ID or purchase ID. The timestamp specifies when an event occurred while the resources are the human or machine resources that perform an activity (Figure 1). This information is taken and used as input data for process modeling, in a discrete-event simulation tool called AnyLogic. The process is broken down into three phases:

Case ID	Start Time	Complete	Activity	Resource	Role
339	31:00.0	23:00.0	Create Pur	Nico Ojenl	Requester
339	34:00.0	40:00.0	Analyze Pu	Maris Free	Requester Manager
339	29:00.0	52:00.0	Amend Pu	Elvira Lore	Requester
339	24:00.0	30:00.0	Analyze Pu	Heinz Guts	Requester Manager
339	36:00.0	38:00.0	Create Rec	Francis Od	Requester Manager
339	34:00.0	58:00.0	Analyze Re	Magdalen	Purchasing Agent
339	50:00.0	03:00.0	Amend Re	Penn Oste	Requester Manager
339	10:00.0	34:00.0	Analyze Re	Francois d	Purchasing Agent
940	31:00.0	08:00.0	Create Pur	Immanuel	Requester

Figure 1: Event log data of a purchasing requisition process

Case study phase I: Modeling of the main process flow using the frequency and performance feature data – the goal of this phase is to use information in process mining frequency and performance features to create a process model of the main process flow. This phase explains how process mining elements such as activities, control flow, and paths are extracted and used to create a simple process model.

Case study phase II: Modeling of Resource units using information from tables in Process mining as input data– this phase is intended to illustrate how resource units' allocation and types in process mining are added to the development of process models.

Case study phase III: Modeling process variations in process mining- the goal of this phase is to zoom in to all other activities in process mining and use this information to model process deviations, loops and conditional paths.

Case study phase I: Modeling of the main process flow using the frequency and performance feature data

After importing the event log data onto the process mining tool, Disco, a process map is generated, from which analysis of the process is done. These analyses are then used to extract useful input data for process modeling. The frequency and performance analysis features in process mining are used.

1) **A process map model of the main flow using frequency feature**: The map generated from event log data in process mining shows the top-down control-flow of the overall process (Figure 2). The map shows a set of activities, which are 18 that the purchasing requisition goes through. The map is set to visualize the frequency of these activities (seen on the right side in figure 2). Frequency displays the number of units moving from one activity to the next (numbers beside the arrows in fig 2) and the number of units being worked on in each activity (numbers inside the rectangles). This information alone suffices as input data to develop the overall basic process model.

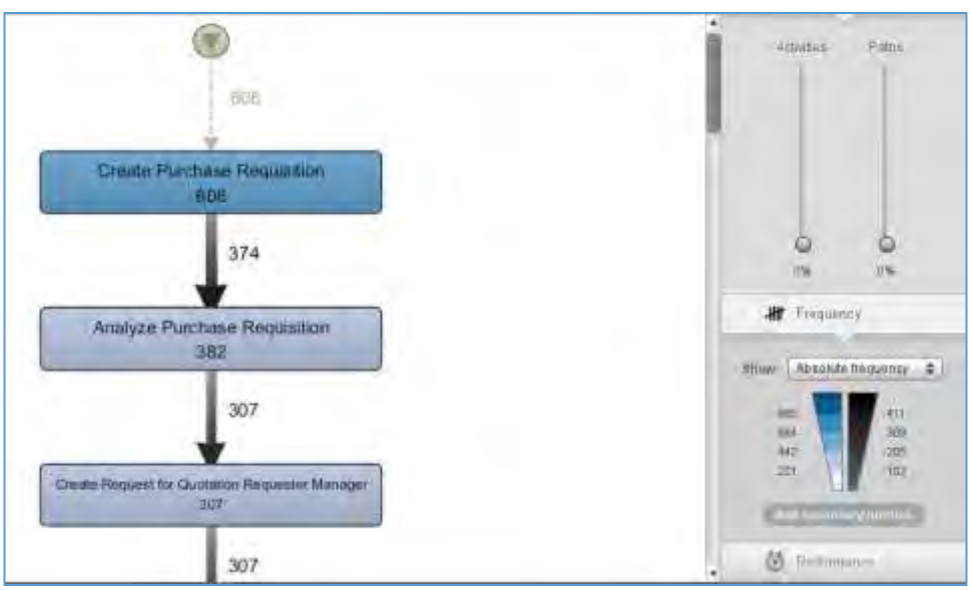


Figure 2: Main process of the purchase requisition on frequency metric in process mining

From Figure 2 above, the process begins with the creation of a purchase requisition. This is represented in process modeling as 'purchasingRequest' agent, the main agent. The main agent is a fundamental concept in process modeling that represents the objects or entities that move through the modeled system. It is the primary entity that performs actions, changes state, and interacts with other objects in the system. To create the purchase requisition in process modeling, the source block is used. This block is the starting point of a process model, it is used to generate model agents. Therefore, since there are observed 608 purchase requisitions created in process mining, the same number is used as input to represent the number of 'purchasingRequest' agents entering the process. The number is entered as 'maximum number of arrivals' in the source block (Figure 3.1). Succeeding activities (Figure 3.2) in process mining are also used as input data for modeling. The processes are modeled using the Queue block in process modeling tool, AnyLogic to represent activities in the rectangles (Figure 3.3). The queue block is a buffer of agents waiting to be accepted by the next block(s) in the process flow. It is therefore the recommended block for modeling activities moving from one rectangle (in process mining), or one queue block (in process modeling) to the next.

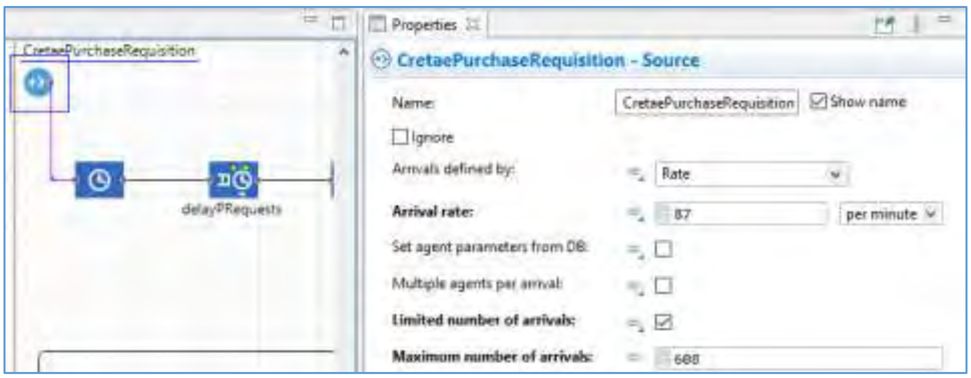


Figure 3.1: Modeling of the main agent in the source block

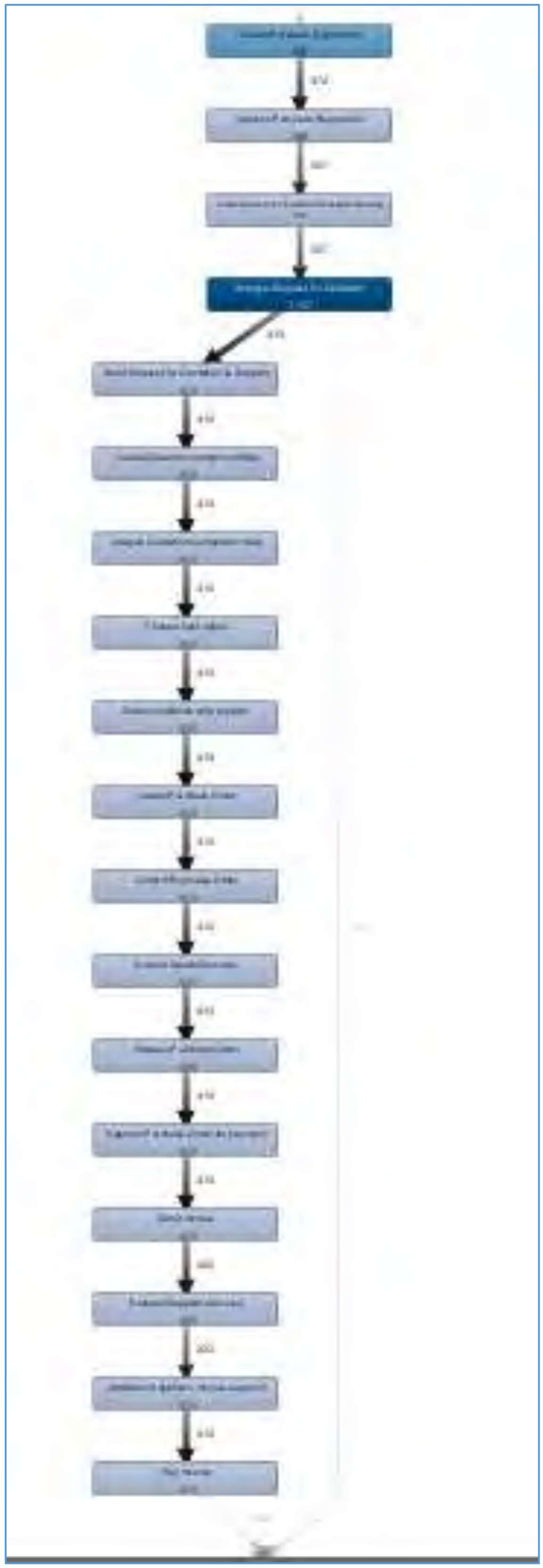


Figure 3.2: Process mining map of the overall process

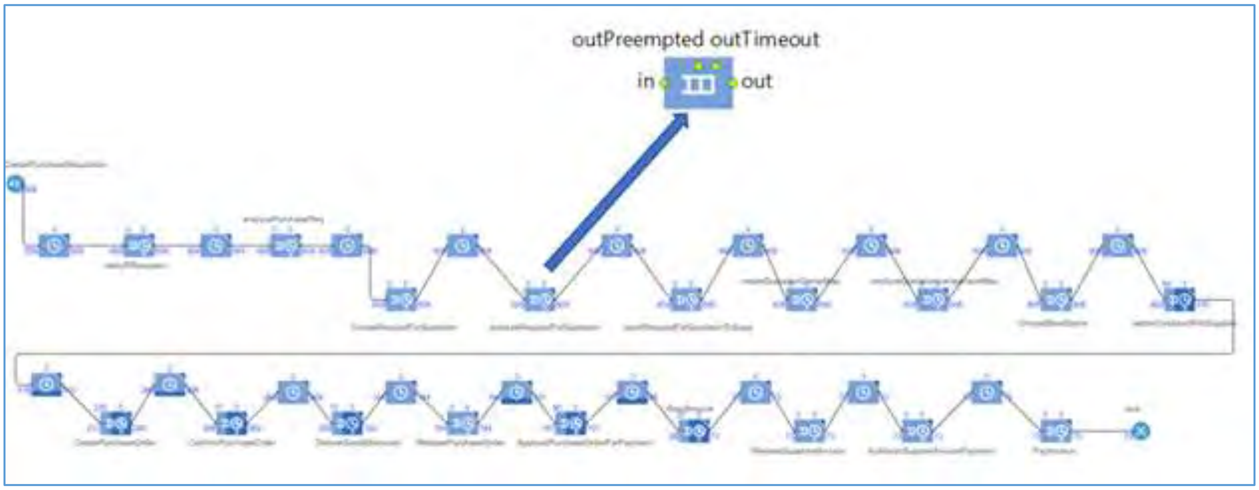


Figure 3.3: Process modeling logic of the overall process

After modeling the frequency of the process, the results obtained are a transition of the process flow from the process mining map in figure 3(b.1) to the process modeling logic of the same process in figure 3(b.2). However, these results only show activities, transitions, and paths which the main agent goes through. The frequency feature lacks the capability to provide any information on the actual performance metrics of the process, such as the time it takes for activities to transition to the next activity. This information is important in identifying delays and other inefficiencies in the process. Therefore, the process map's performance feature is applied to display key indicators of process efficiency such as cycle time, lead time, and processing time.

2) **A process map of the main flow using performance feature:** After applying the performance feature to the process map, more detailed and granular insights into process performance are obtained. This information helps in identifying bottlenecks, inefficiencies, and areas for improvement. These are key insights that Managers who use process analysis techniques look for. Therefore, it is important to add this information as input data for developing process models. Figure 4 shows the visualization of the process map when the performance feature is applied.

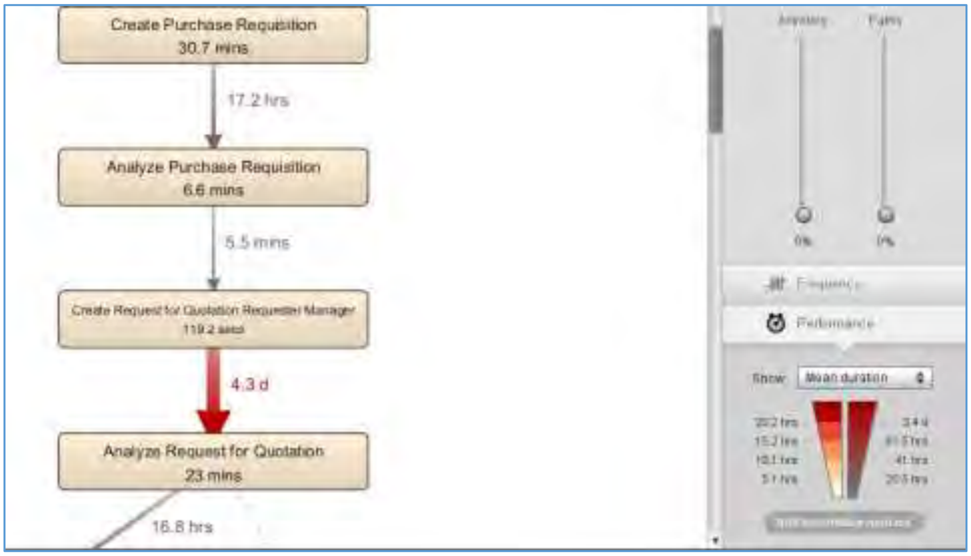


Figure 4: Main process of the purchase requisition on performance metric in process mining

The numbers inside the rectangles represent time that was taken by a certain resource to perform the activity, while the number beside the arrow indicates the time taken for the main agent (the purchasing requisition) to move from one activity to the next. These numbers are used as input data in process modeling, where the Delay block in process modeling is used to represent the time in which the agent takes to move from one activity to another. The delay block can delay agents for a given amount of time. It evaluates the agent's delay time dynamically, may be stochastic and depends on the agent and any other conditions. It also allows for multiple agents (up to the given Delay capacity) to be delayed simultaneously and independently. For example, on the process map, the agent took 17.2 hours to move from the first activity 'Create Purchase Requisition' to the second activity, 'Analyze Purchase Requisition'. This is represented in the delay block's parameter (Figure 5.1), where the delay time is set to 17.2 hours. This will make the transition of the agent from the first activity to the second, to take exactly this amount of time when the model is run (figure 5.2). This performance or behavior can be observed in the developer panel (on the right side in figure 5.2) which provides access to the model data during the model run time.

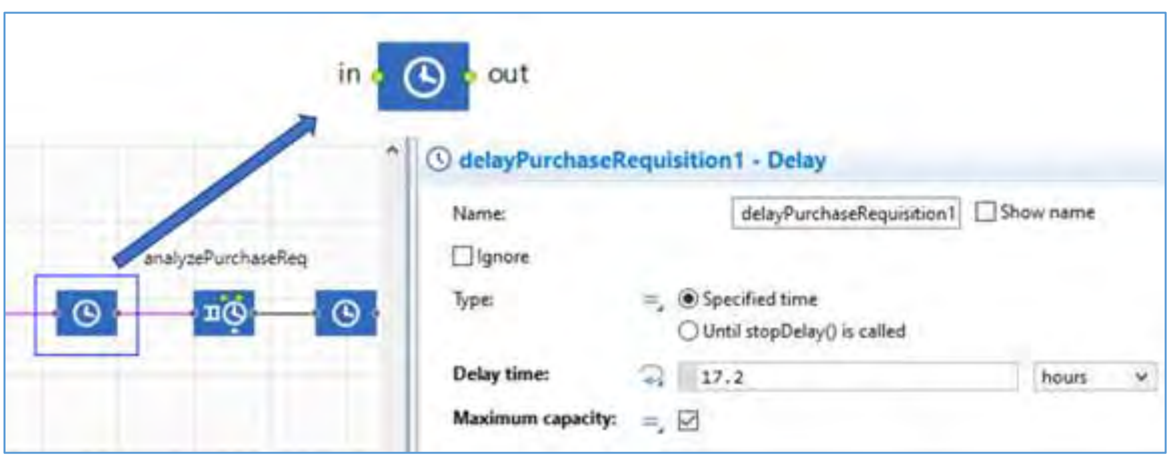


Figure 5.1: The delay parameter in process modeling, representing the time taken from one activity to the next in process mining

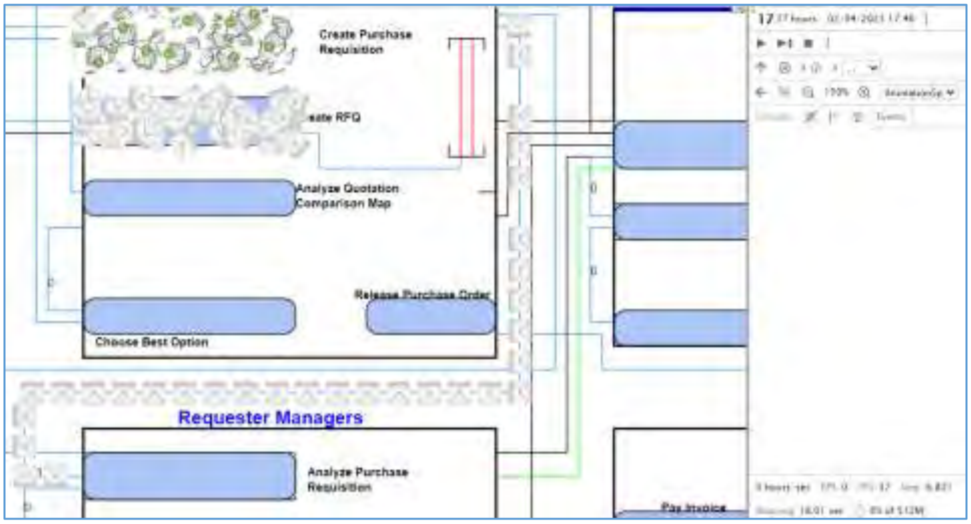


Figure 5.2: Process modeling output showing transition from activity ‘Create Purchase Requisition’ to activity, ‘Analyze Purchase Requisition' taking 17.2 hours

Additionally, the process mining data also shows the duration of time that each activity’s resource agent took to work on the purchase requisition (the numbers in the rectangular shapes). For example, activity three, 'Create Request for Quotation Requester Manager’ took 119.2 seconds (about 2 minutes) to be processed by a particular resource agent. This information is used as input in process modeling, where the Service block is used to model this process. The function of the service block is to delay the agent, seize a given number of resource units, and release the seized units. Since there are no resource units up to this point, the seize block is used to delay the agent. The 119.2 seconds (about 2 minutes) was used in the delay time parameter area of the service block (Figure 6), to delay the agent in activity three, for this period.

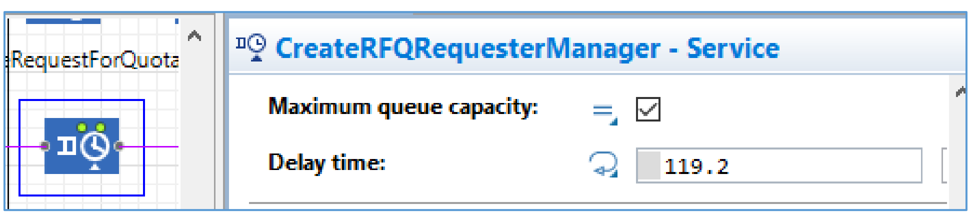


Figure 6: The Service block parameter in process modeling representing the time an agent takes in an activity

After running the model, activity 3 was observed in the developer panel to take 119.2 seconds (about 2 minutes) before moving to the next block. After developing the main process flow’s activity frequency and performance, the main flow is seen to lack insight into the behavior of individual process participants, or the interactions between them. This is particularly an important aspect of any business process. As mentioned by Doomun & Vunka Jungum, (2008) that a business process is a series of connected entities that use the organization’s resources to produce a specified output for a particular customer or market. Therefore, the process mining tool was further analyzed to identify the resources, and their relationships.

Case study phase II: Modeling of Resource units using information from tables in Process mining as input data

Among the output information displayed in Disco, process mining tool, 'cases' are one of them. They are represented as rows in the event log (Figure 7). Each row contains information about a specific event in the process, such as the activity that was performed, the time at which it occurred, the resource that executed it, the date the activity was executed, and duration it took for each activity (Rozinat et al., 2009).

Since the model with specific activities and times has been created, the focus of this phase is to implement the resource agents that oversaw these activities. The following resource agents were identified in the 'cases' section of process mining: Requesters, Request managers, Purchasing agents, Financial managers, and Suppliers. 608 cases were also observed, but only case 335 was selected to represent the main flow. Case 335 was seen to have a similar number of activities as in the main process map. Using the table function under cases in process mining, it is possible to identify the major roles of the resource agents and the activities they were involved in at a particular point in time.

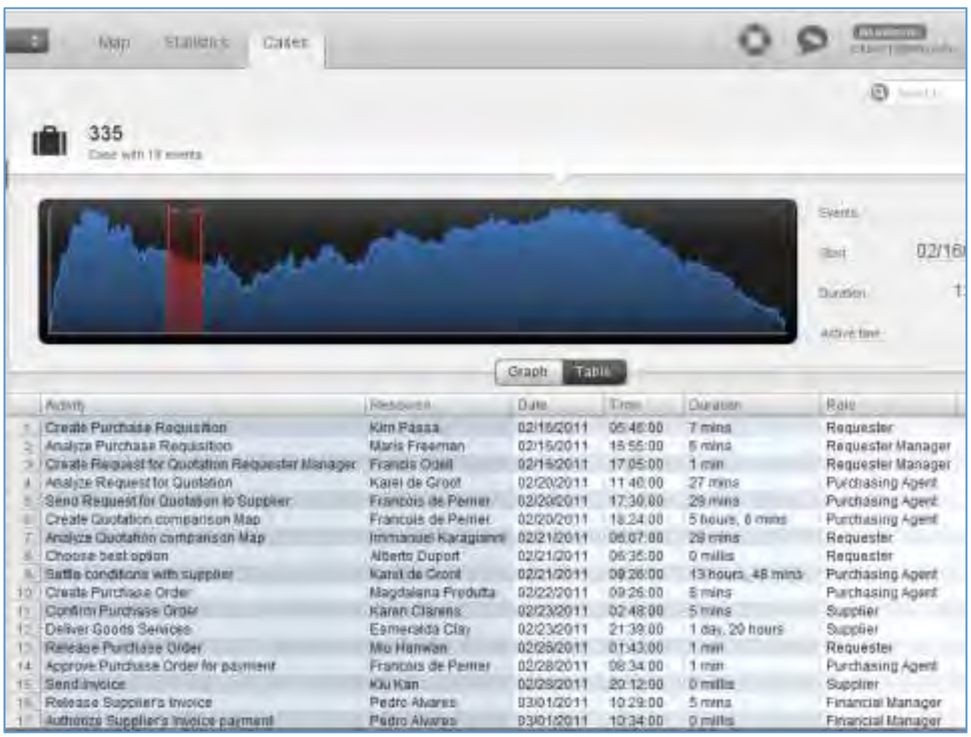


Figure 7: Resource agents in process mining

This information was then used to create a process model, containing resources. In the process modeling tool, the Resource Pool block was used. It is used to define a set of resource units that can be seized and released by agents. It can be of three types: 1) Static resources which are bound to a particular location, 2) Moving resources which can move on their own, they can represent staff, vehicles, etc. And 3) Portable resources which can be moved by agents or by moving resources. In this case study, the resource agents' type was static as they did not require movement to carry out a task. The resource pool block can define the capacity of the resource agents in the process. In this case, requesters in process mining were counted to have carried out activities 4 times, the request managers 3 times, Purchasing agents 3 times, financial managers 3 times, and suppliers 3 times. These numbers were then used in the process modeling to represent the capacity of the resource agents using parameters (Figure 8)

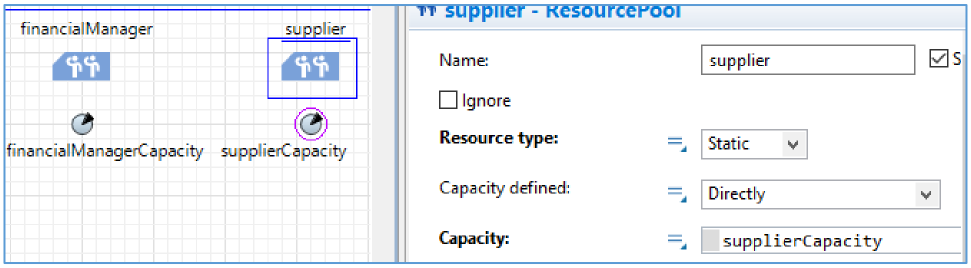


Figure 8: Resource pool block in process modeling

After introducing the resource agents in the process model, the resources were then connected to the activities they carried out. This is done in the service block's resource pool parameter from where the resources are requested for a specific activity. After modeling the resources, their utilization was measured using the key performance indicators in process modeling to help identify areas where resources are being over-utilized or under-utilized. Table 1 shows the analysis results of resource utilization after the model is run for one month. This information can be used to optimize resource usage and improve the overall efficiency of the system (Abohamad et al.2017).

Table 1: Resource Utilization Key Performance Indicators (KPI) from Process Modeling

Resource	**%Utilization**
Requester	34%
Requester Manager	97%
Purchasing Agent	70%
Financial Manager	10%
Supplier	10%

After completion of the model with resources included, it was important to realize that modeling only one case out of 608 cases does not really represent the overall process reality as variations in processes are inevitable. The next phase attempts to bridge this gap by modeling subprocesses identified in process mining.

Case study phase III: Modeling process variations in process mining

Real processes are complicated, and often complex such that it is not possible to look at every detail at once (Rozinat et al., 2009). From the process mining information, there are 98 process variations uncovered. These are specific instances or occurrences of the process when activities were executed multiple times. Process variants exhibit different variations in terms of the activities performed, the sequence in which they are performed, and the time taken to complete them. The variants are captured as different process instances in process mining (Figure 9.1)

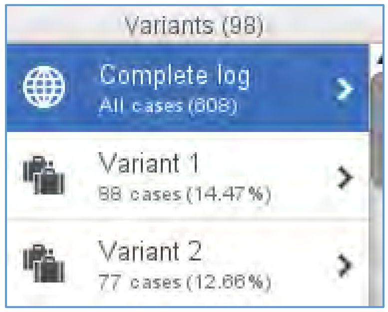


Figure 9.1: 98 Variants uncovered from event log data in process mining

Process mining algorithms, however, have the capability to choose the level of detail to look at processes (Rozinat et al., 2009). Using the paths and activities features in process mining, it is possible to identify the variant that represents the overall process flow map. When activities and paths sliders are put to their lowest point, 0%, only activities performed in the most frequent process variant are observed, which results in visualization of the main flow of the process. When moved to the highest point, 100%, every activity and path ever performed is visualized. (Figure 9.2)

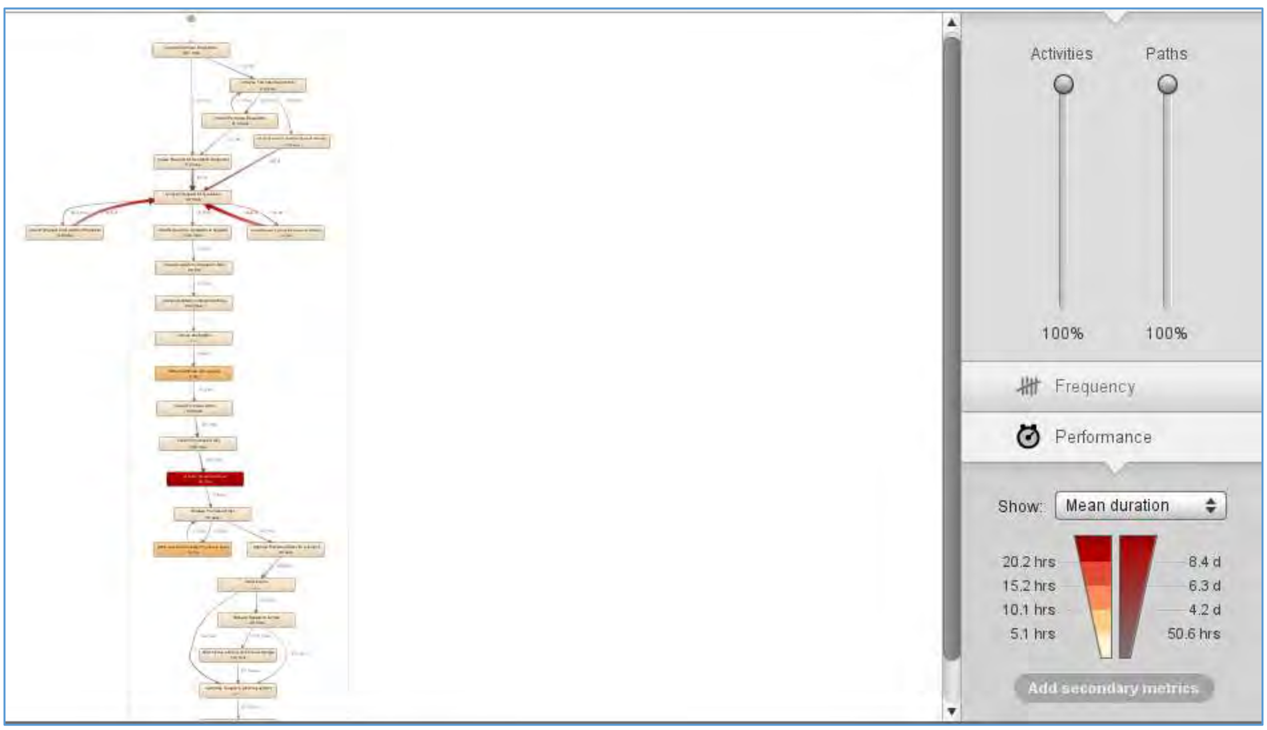


Figure 9.2: Sub process of all the activities and paths

The results in the visualization show 24 activities done. This additional information is then used as input data to create an advanced process model with more routes and sub processes. A variant that represents 24 activities is chosen at random. It is represented in table format in process mining (Figure 9.3). This variant contains information such as activities, resources, time and duration.

	Activity	Resource	Date	Time	Duration	Role
8	Analyze Request for Quotation	Karel de Groot	07/22/2011	00:09:00	18 mins	Purchasing Agent
9	Send Request for Quotation to Supplier	Magdalena Predutta	07/22/2011	13:46:00	22 mins	Purchasing Agent
10	Create Quotation comparison Map	Francois de Perner	07/22/2011	18:07:00	2 hours, 28 mins	Purchasing Agent
11	Analyze Quotation comparison Map	Anna Kaufmann	07/23/2011	03:27:00	27 mins	Requester
12	Choose best option	Fjodor Kowalski	07/23/2011	03:54:00	0 millis	Requester
13	Settle conditions with supplier	Magdalena Predutta	07/23/2011	17:00:00	14 hours, 28 mins	Purchasing Agent
14	Create Purchase Order	Magdalena Predutta	07/24/2011	10:00:00	10 mins	Purchasing Agent
15	Confirm Purchase Order	Carmen Finacse	07/24/2011	22:16:00	7 mins	Supplier
16	Deliver Goods Services	Carmen Finacse	07/26/2011	11:05:00	1 day, 16 hours	Supplier
17	Release Purchase Order	Esmana Liubiata	07/28/2011	10:23:00	1 min	Requester
18	Settle dispute with supplier Purchasing Agent	Francois de Perner	07/28/2011	13:07:00	9 hours	Purchasing Agent
19	Release Purchase Order	Anna Olwada	07/29/2011	06:37:00	1 min	Requester
20	Approve Purchase Order for payment	Karel de Groot	07/29/2011	15:32:00	1 min	Purchasing Agent
21	Send invoice	Sean Manney	08/01/2011	01:11:00	0 millis	Supplier
22	Release Supplier's Invoice	Karalda Nimwada	08/01/2011	15:23:00	7 mins	Financial Manager
23	Authorize Supplier's invoice payment	Pedro Alvares	08/01/2011	15:30:00	0 millis	Financial Manager
24	Pay invoice	Karalda Nimwada	08/01/2011	16:31:00	13 mins	Financial Manager

Figure 9.3: Table showing variant details that have 24 activities

When observed closely in the process map, some activities are seen to be branching off from activities in the main path. For example, the activity ‘Amend Request for Quotation Requester’ (Figure 9.4) is one of the activities branching off from the main process activity, ‘Analyze Request for Quotation’. It takes 50.2 minutes from the main activity and 8.5 days back. Other examples of such activities are shown in Figures 9.4 and 9.5. All the information about these activities is represented in figure 9c above.

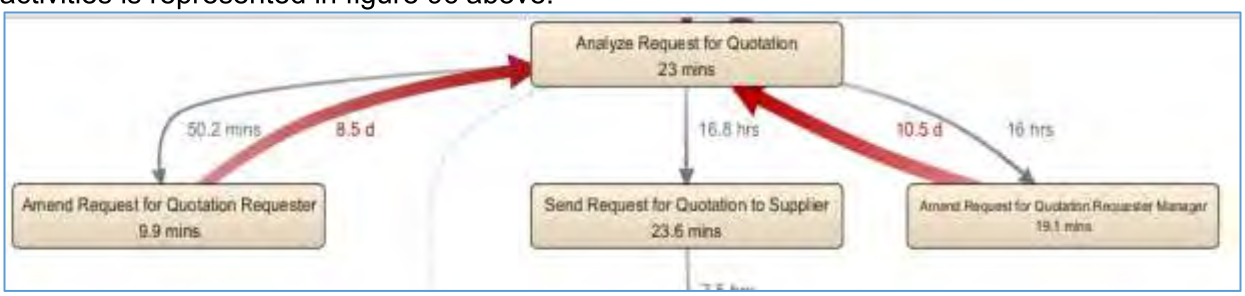


Figure 9.4 Activities branching off from the main process

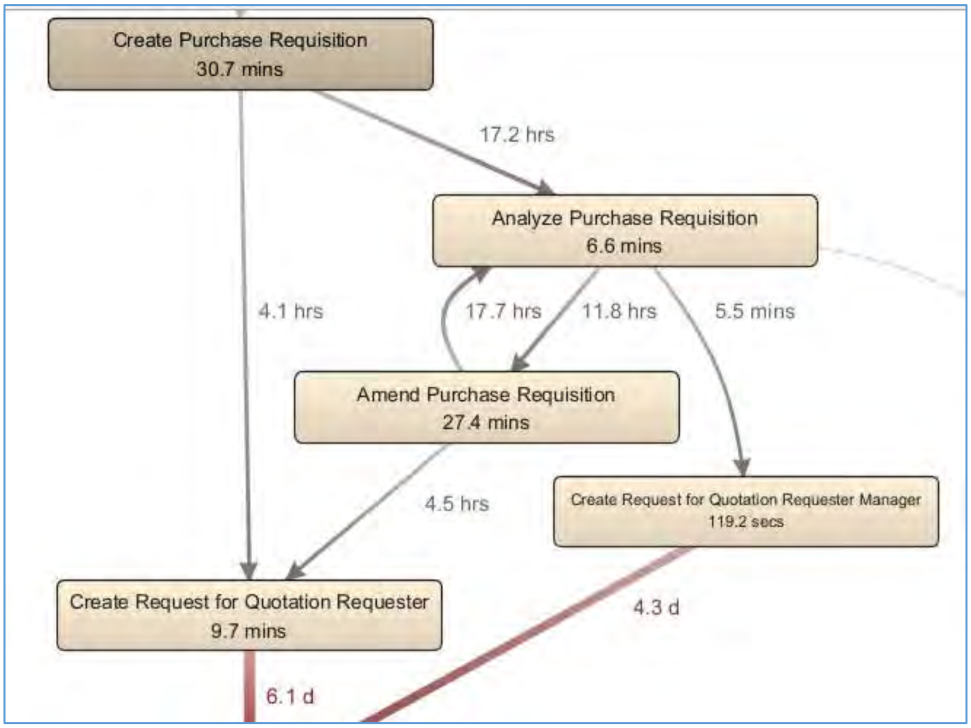


Figure 9.5 Additional activities branching off from the main process

To model such activities, an additional block in process modeling is introduced, the select output block (Figure 10.1). It is typically used when there are multiple output paths available from a process, and the process model must decide which path to take. It is used here as a tool for making decisions on where the activities are routed to. Since this information is provided in process mining, the select output is then used at all instances where the branching of activities is taking place.

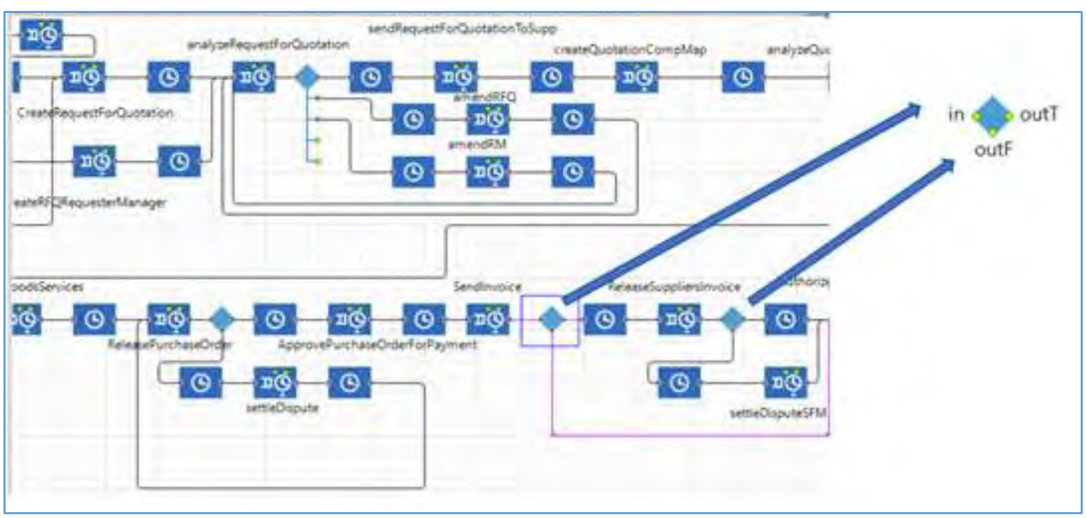


Figure 10.1: Subprocesses representation in process modeling

One challenge with modeling this information however is that, usually there is a reason for activities to move back and forth between resource agents. This kind of information is obtained from interviewing process owners or managers. However, since there is no record for reasons why the activities are branching off to different resource agents, the model used the select output default probability to branch off these activities. Had there been information on why activities branch off, the select output block would have been set to route each individual activity to its succeeding activity based on a condition created in java programming. After modeling the subprocesses, the duration of time taken from the start of the process up to the end was measured and found to be 1004.63 (Figure 10.2). To measure duration in process modeling, 'TimeMeasureStart' as well as 'TimeMeasureEnd' blocks are used (Figure 10.3). The blocks measure the time the agents spend between activities by remembering the time when an agent went through the 'TimeMeasureStart'. Later, when the agent goes through TimeMeasureEnd block, the time the agent spent between these two "marker" blocks is measured.

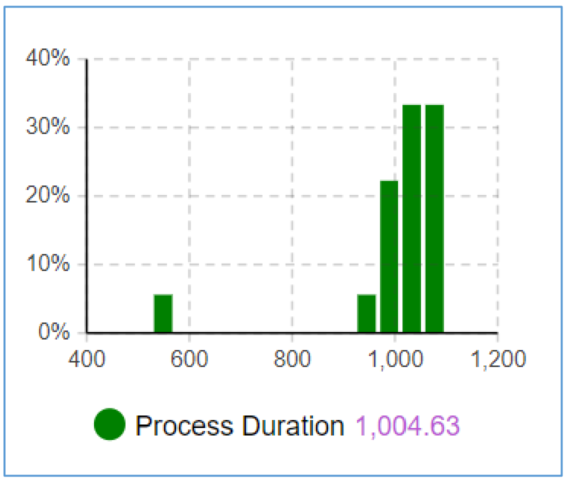


Figure 10.2: Process duration from start to finish

Figure 10.3: Time Measure Start block in process modeling

This information provides valuable insight into the performance and efficiency of the system. By measuring the time taken to complete a process or a set of processes, we can identify bottlenecks, inefficiencies, and areas for improvement. This information can then be used to optimize the system, increase productivity, and reduce costs. For example, in the process mining output, an observation is made where activity 'Deliver goods Services' is highlighted in red, to show that this activity is a pain point in the process, a main bottleneck due to the length of time it takes to execute this process (Figure 10.4). This issue is also observed after the process is modeled, and the output results show overcrowding at this activity (Figure 10.5).

Figure 10.4: Illustration of the problems and inefficiencies in process mining.

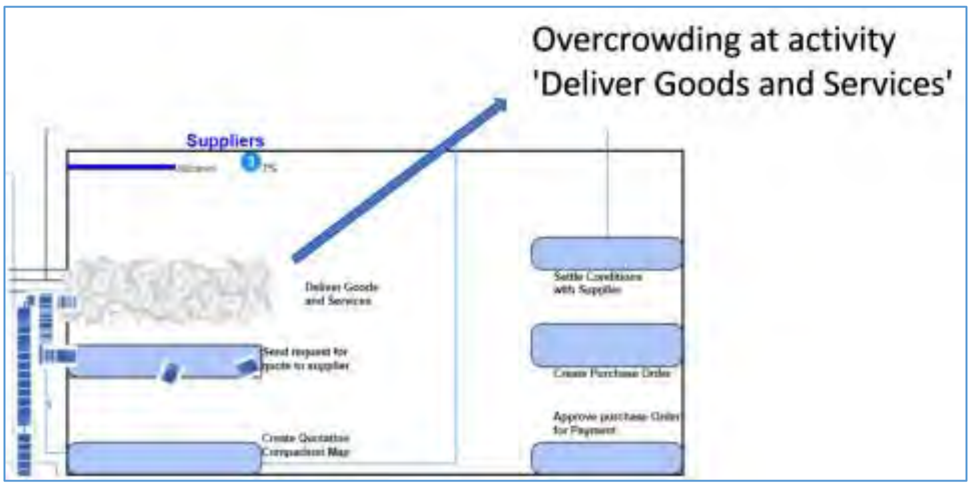


Figure 10.5: Illustration of the problems and inefficiencies in process modeling.

Using process modeling, a what-if scenario is created to mitigate this issue. An improvement suggestion is made to omit activity ‘Amend Request for Quotation Requester Manager’, as it is seen to take 16 hours and 10.5 days (about 1 and a half weeks) to move back and forth from activity ‘Analyze Request for Quotation’ (Figure 10.6).

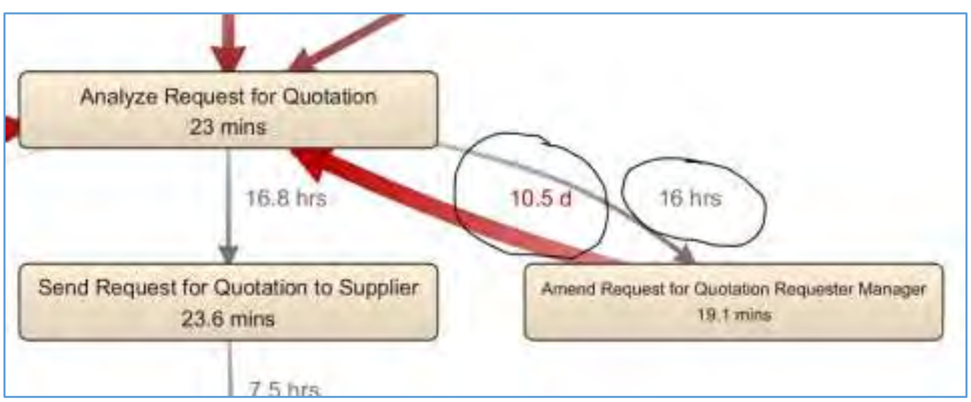


Figure 10.6: An activity assumed to cause the root problem

Another improvement suggestion is to assign the task ‘Amend Request for Quotation ‘to the Requester only as the task is being performed by the Requester Manager Resource as well, causing a waste in the utilization of resources. The utilization for the Requester had been observed to be 34% initially, while that of the Requester Manager was 97%. After the application of the suggested improvements, the results show that utilization of the Requester manager improved from 97% to 70% (Figure 10.7) while the cycle time of the process is significantly reduced from 1082.86 to 844.67 (figure 10.8 and 10.9).

Figure 10.7: Observed change in resource utilization of the Requester Manager

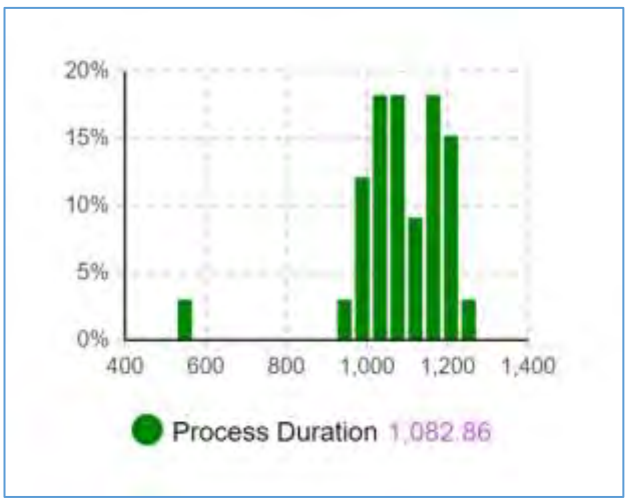


Figure 10.8: Process Duration without improvement suggestions

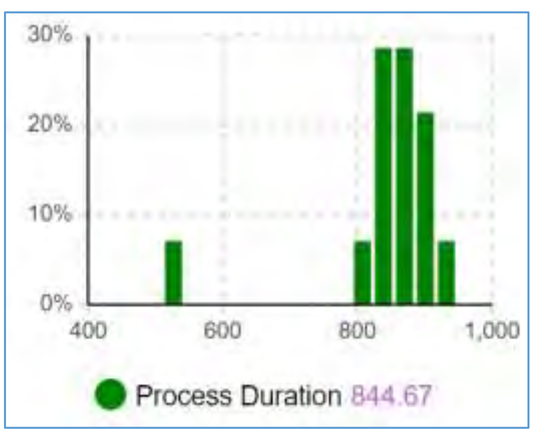


Figure 10.9: Process Duration with improvement suggestions

After applying these improvement suggestions, process modeling enabled the evaluation of the impact of adding or removing tasks to resources or changing the workflow. Making process modeling technique more optimizable and capable of reducing the cycle time, increasing throughput and improving the quality of the process.

LESSONS LEARNED

In the attempts to apply process mining analytics information as input data for more accurate process models, the major lessons learned are summarized as follows:

1. Process mining and process modeling techniques are complementary. They provide different perspectives on the same process where process mining provides visualization of the entire process information about the process flow, timestamps, activity names, and resource assignments.
2. Process modeling, however, uses the same information to build models that reflect the actual process flow.
3. Process mining can easily help identify areas of the process that need improvement
4. While process modeling can be used to test potential changes to the process before they are implemented.
5. Process modeling also provides more information on process elements like resource utilization and duration. This is achievable through its key performance indicators that allow for easy analysis and measurement of process performance.
6. With process modeling capability to create what-if scenarios, continuous improvement of the system is achieved.

By combining the insights gained from both approaches, organizations can make data-driven decisions to improve their processes and achieve better performance outcomes.

CONCLUSION AND LIMITATIONS

In this paper, the process mining analytics was instrumental in uncovering the system’s real process flow and related information, which would not have been acquired through interviews and subjective observations. Extracting such real process information from the process mining gives tangible insight on the major activities, transitions, and paths that the main agent goes through. This, however, presents a limitation where information on the process's performance metrics is not represented. To fill this gap, the performance feature is used which uncovers key information such as cycle time, lead time, and processing time. While creating a process model with this information, it is also realized that the main flow lacks insight into the behavior of individual process participants, or the interactions between them. This led to the use of the roles feature in process mining which has information on human resources that perform certain activities. Modeling this information provided more insight into the utilization of such resources using the key performance indicators feature in process modeling. It was also clear that processes are complex. The process mining information had different activities and sequences that were performed at various times. This is captured in process mining in the form of variants. There were 98 variants uncovered from this case study, from which only one that stood for the entire flow of the process was modeled. Selecting one out of 98 variants presents a challenge as the process model might not accurately represent the real process.

Nevertheless, the results from modeling the activities resulted in the same output seen in process mining flow, where an event of a bottleneck that was visualized in red in process mining, had the same behavior where the same instance was seen to overcrowd in process modeling. This acted as validation that process mining data can be used as input data for process modeling. Additionally, by comparing the results of the process model with the actual process flow in process mining, users can find any discrepancies and make necessary adjustments to the model. This was an advantage realized from using process modeling. It's capability to edit the model and to create what-if scenarios allowing for experimentations. This led to a suggestion for a solution to the main process flow. The solution was then implemented, and actual improvement results were obtained. However, for any major changes in the process flow, confirmation is needed from process managers or process owners. Modelers need to conduct interviews or meetings to discuss the changes before implementing them. This was not captured in the model as it relied on information from process mining alone and the modeler's assumption.

Another limitation was that once processes were modeled, there was no way to confirm that it produced the same information as that in process mining. However, the model's behavior, the same as that of the process mining output, acted as validation. In conclusion, by combining these two approaches, organizations can find process variations and inefficiencies and then use process modeling to design and implement more efficient processes. This approach has been successfully applied in this paper. The overall combination of process mining and process modeling is seen to be powerful for improving organizational processes and achieving greater efficiency and effectiveness. Further research is needed to explore process mining techniques which lack the capability to create hypothetical scenarios and observe different outcomes. Future work is recommended to study how process mining predictive capability can be improved using process modeling.

REFERENCES

Aguirre, S., Parra, C., & Alvarado, J. (2013). Combination of process mining and simulation techniques for business process redesign: A methodological approach. *Lecture Notes in Business Information Processing*, 24–43. https://doi.org/10.1007/978-3-642-40919-6_2

Augusto, V., Xie, X., Prodel, M., Jouaneton, B., & Lamarsalle, L. (2016). Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation. 2016 Winter Simulation Conference (WSC). https://doi.org/10.1109/wsc.2016.7822256

Banks, J. (1998). *Handbook of Simulation: Principles, methodology, advances, applications and practice*. Wiley.

Brailsford, S. C. (2014). *Discrete-event simulation and system dynamics for Management Decision making*. Wiley.

Curtis, B., Kellner, M. I., & Over, J. (1992). Process modeling. *Communications of ACM*, *35*(9), 75–90. https://doi.org/10.1145/130994.130998

Doomun, R., & Vunka Jungum, N. (2008). Business process modelling, simulation and reengineering: Call Centers. *Business Process Management Journal*, *14*(6), 838–848. https://doi.org/10.1108/14637150810916017

Greasley, A. (2003). Using business-process simulation within a business-process reengineering approach. *Business Process Management Journal*, *9*(4), 408–420. https://doi.org/10.1108/14637150310484481

Günther, C. W., & van der Aalst, W. M. (2007). Fuzzy mining – adaptive process simplification based on multi-perspective metrics. *Lecture Notes in Computer Science*, 328–343. https://doi.org/10.1007/978-3-540-75183-0_24

Jadrić, M., Pašalić, I. N., & Ćukušić, M. (2020). Process mining contributions to discrete-event simulation modelling. *Business Systems Research Journal*, *11*(2), 51–72. https://doi.org/10.2478/bsrj-2020-0015

Mahmood, I., Kausar, T., Sarjoughian, H. S., Malik, A. W., & Riaz, N. (2019). Integrated modeling, simulation, and Analysis Framework for Engineering Complex Systems. *IEEE (Institute of Electrical & Electronic Engineers) (Institute of Electrical & Electronic Engineers) (Institute of Electrical & Electronic Engineers) (Institute of Electrical & Electronic Engineers) (Institute of Electrical & Electronic Engineers) (Institute of Electrical & Electronic Engineers) Access*, *7*, 67497–67514. https://doi.org/10.1109/access.2019.2917652

Martin, N., Depaire, B., & Caris, A. (2015). The use of process mining in business process simulation model construction. Business & Information Systems Engineering, 58(1), 73–87. https://doi.org/10.1007/s12599-015-0410-4

Măruşter, L., & van Beest, N. R. (2009). Redesigning business processes: A methodology based on simulation and process mining techniques. Knowledge and Information Systems, 21(3), 267–297. https://doi.org/10.1007/s10115-009-0224-0

Robinson, S., Radnor, Z. J., Burgess, N., & Worthington, C. (2012). Sim Lean: Utilising simulation in the implementation of Lean in Healthcare. *European Journal of Operational Research*, *219*(1), 188–197. https://doi.org/10.1016/j.ejor.2011.12.029

Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2009). Discovering simulation models. *Information Systems*, *34*(3), 305–327. https://doi.org/10.1016/j.is.2008.09.002

van der Aalst, W., Weijters, T., & Maruster, L. (2004). Workflow mining: Discovering process models from event logs. IEEE Transactions on Knowledge and Data Engineering, 16(9), 1128–1142. https://doi.org/10.1109/tkde.2004.47

Wang, Y., Zacharewicz, G., Traoré, M. K., & Chen, D. (2018). An integrative approach to Simulation Model Discovery: Combining system theory, process mining and Fuzzy Logic. *Journal of Intelligent & Fuzzy Systems*, *34*(1), 477–490. https://doi.org/10.3233/jifs-17403

DECISION SCIENCES INSTITUTE

Using education analytics to predict at-risk students in the university: Case Study

Ma, Nang Laik
Singapore University of Social Sciences
Email: nlma@suss.edu.sg

Chia, Sook May Ivy
Singapore University of Social Sciences
Email: jivychiasm@suss.edu.sg

ABSTRACT

Analytics for Education and Learning has become very popular in recent years. In this paper, the authors created a use case to allow students to explore the data and develop three predictive models to predict students' CGPA to identify if the student is at-risk academically. Finally, students will choose the best model based on the mean square error (MSE). At the end of the case, students will recommend a few intervention methods to improve their academic performance and thus improve the university's reputation and increase the students' graduation rate.

KEYWORDS: education analytics, data analysis, predictive model, at-risk students, intervention methods

1. INTRODUCTION

Educational institutional personnel must make strategic, tactical, and teaching decisions through data analytics. Education institutions collected many students' data from student demographics, a learning management system that keeps track of students learning activity, grades, student e-portfolio, and digital library activities. Due to COVID-19, as most of the classes are moving online, there is no lack of digital traces of students available, and the institution may lack resources to use and interpret these learning activity data to highlight any issues students may face. As many student activity data are collected within the education institutions, senior management are expected to read and analyze institutional data. In the authors' university, a business analytics unit (BAU) has been set up to look at institutional analytics, providing dashboard and predictive analytics results to assist lecturers and administrators in making sound decisions based on actionable insights. Thus, big data analytics across the different educational institutions and integrating the learning analytics course within the adult learning degree program at the authors' university becomes more crucial.

Analytics for Education and learning analytics enables students to understand how raw data could and should be handled so that useful information can be extracted from the data. A knowledge society presents unique and fundamental questions of what data means and how to use it responsibly and well. In particular, the recent advent of big data and analytics has magnified the question, and creates both opportunity and conundrum for organisations and institutions, there is increasing pressure to make efficient decisions at the institution level by looking at the data and identifying at-risk students. An at-risk student in our university definition is a student whose

Cumulative Grade Point Average (CGPA) is less than 2.0 out of 4 for two consecutive terms. The university believed that these students have a high probability of dropping out of the university if no action is taken. A student must achieve a minimum CGPA of 2.0 upon graduation to graduate from our university.

To identify at-risk student is one of the major issues in most universities around the world as it affects the university reputation, and financial status. Thus, institutions are willing to spend millions of dollar to develop analytics solution to predictive students who are at-risk and require temporary or ongoing intervention to succeed academically. The institutions will develop an early alert system using the students learning management system, historical academics grade, and developing a model to predict the students' academic performance to help students graduate on time and be more successful. One of the assumptions of these systems is that data collected from the student management system is a proxy of how well they are learning, and the future students' performance will be similar to the past students' performance.

This case study developed by the authors here is based on the authors' experiences of running the programme in the university. Due to the data privacy issues, the data generated does not include any sensitive data of students' ID. However, the techniques students applied to solve this use case will enable them to solve real-world problems. The case study has targeted undergraduate and post-graduate students pursuing a degree in data analytics or decision science subjects related to education. There is no pre-requisite to the course, and the instructors will use Excel and the IBM SPSS modeler or other data mining software such as SAS or Python to teach this course.

The case study requires exploratory data analysis, understanding various predictive models, comparing the model performance, and sharing valuable insights from the data with the audience. The authors would like to recommend that the instructors who use the case study give students ample time to go through the data and solve the problem. Students should be encouraged to look at the model in detail and not just run the model and share some graphs within any recommended actions.

The rest of the paper is organized as follows: Section 2 is the literature review, Section 3 explains the business problem and explores the data. Next, Section 4 shows the development of three predictive model and summarizes the results. Finally, the paper ends with a broad discussion on intervention method and suggestion to use the model results.

2. LITERATURE REVIEW

Institutions worldwide are looking to use the students' management system data to track and analyze results across different student cohorts and make better decisions to improve the educational outcomes (Aldowah, Al-Samarraie & Wan Mohamad 2019, and Chiappe & Rodriguez 2017).

(Barrows 1996) explained the educational objectives of problem-based learning (PBL) and encouraged the management to give the faculty full support. The support will help the faculty change the established curriculum to PBL to see the impacts on students learning capability and independent thinking. In the meantime, problem-based learning (PBL) has been the main focus of developing learning analytics courses, according to Ma & Chia (2020). The authors used the framework to create the master-level course and received good feedback.

Junco and Clem (2015) developed an early warning system to identify at-risk students using digital course reading and engagement. The data source comes from the digital reading platform, and proxy to engagement as the number of pages read, reading sessions, days interacted, time spent, and the number of highlights, bookmarks, and notes. Additional data sources come from previous GPA, course grades, and demographic information. The result showed that the engagement and number of days' students spent on reading were strong predictors of student performance. One challenge would be that if students printed book pages, this hinders the data collection where low engagement would be measured. The system developed allows the educator to provide early intervention and teaching and learning strategies such as reaching out to contact the student or coaching to improve student performance.

3. BUSINESS PROBLEM AND DATA EXPLORATION

Students are given the case at the beginning of the class, including the problem description and a dataset. They are encouraged to work in a team of three to four. They need to analyze the data and propose a data analytics solution to identify at-risk students for a given dataset at the end of the class. The duration of the class is about 4 hours. In the first half an hour, the instructor goes through the case, and the brainstorming session continues within a team. They discussed some of the reasons for at-risk students: health issues, change in financial status (e.g., a father was retrenched and thus the student needed to work part-time to support the family), emotional problems, family violence, and more. The problem analysis in this case study started with understanding the problem and making appropriate assumptions. After initial "brainstorming," students should be able to list the critical inputs of the model are:

- CGPA is the Cumulative Grade Point Average, and from the data, it is between 0 – 4 points. CGPA is computed based on the score/grade of each subject for all the previous semesters. The higher the CGPA, the better the academic performance of the students. The highest CGPA is 4.0.
- The test scores are numerical scores from various subjects, in this case, we have identified five subjects that are good predictors of CGPA. The score is between 0 – 100 marks.

In the university, CGPA is a good measure of students' academic performance in the university. Students, who do not meet the minimum academic performance, CGPA of less than 2.0 for two consecutive terms, will be asked to leave the university to maintain its academic quality, the university standard, and reputation. Thus, the school management teams need to regularly look at the students' CGPA at the end of each semester. Assume that the students are part of the analytics team and a task to develop a predictive model to determine the crucial variables that are good predictors for the target or dependent variable, CGPA. Out of 40 courses students need to graduate, five test scores are selected, and 2000 students' records and their CGPA are recorded. Students must perform some initial data analysis, present their findings, and propose analytics solutions as proof of concept to the university management team, including people from the business and IT divisions.

To achieve this, students should do the following tasks (but not restricted to):

a. First, perform exploratory data analysis on the data given.
b. Then, develop three predictive models to predict the students' CGPA based on the inputs.
c. Describe each model in detail and state the essential parameters.
d. Analyze the pros and cons of the chosen model and model limitations
e. Compare and validate the models developed above.
f. Give a recommendation and suggest the most suitable for implementation.

Table 1: Student data

Description	Data Field
Student ID	Categorical
gender	0 - Male, 1 - Female
age	Numeric
Score 1	Subject 1 final score (0 - 100)
Score 2	Subject 2 final score (0 - 100)
Score 3	Subject 3 final score (0 - 100)
Score 4	Subject 4 final score (0 - 100)
Score 5	Subject 5 final score (0 - 100)
CGPA (Target)	CGPA (0 - 4)

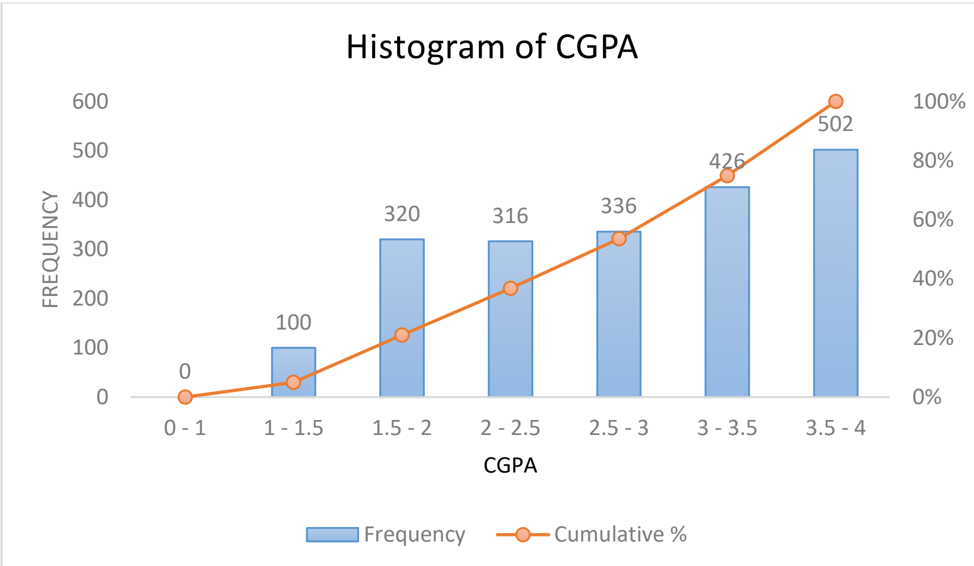
There are 2000 students' records from the past five years, and there is no missing data. The test scores are between 0 and 100, and the CGPA ranges between 0 and 4. Table 2 shows the summary statistics of the data. Since we are interested in CGPA as the target variable, the mean CGPA is 2.79 with a standard deviation of 0.76, the median is 2.89, and the mode is 3.56, which means that most students have good academic performance. The minimum CGPA is 1.23, and the maximum is 3.99.

Next, we should look at the input variables. The gender ratio is 66% male and 34% female. The student's age is from 18 to 25, the average is 21.48. The average score5 is the lowest among all the input, which is 60.22, and the highest average score is score4, which is 74.82. Score1, 2, and 3 have an average of 68.41, 65.98, and 65.4.

Table 2: Summary statistics of students' data

	age	score 1	score 2	score 3	score 4	score 5	CGPA
Mean	21.48	68.41	65.98	65.40	74.82	60.22	2.79
Median	21.00	70.00	65.00	68.00	78.00	60.00	2.89
Mode	19.00	75.00	75.00	74.00	88.00	45.00	3.56
Standard Deviation	2.32	9.37	12.20	13.39	14.79	13.51	0.76
Sample Variance	5.37	87.89	148.87	179.27	218.78	182.39	0.58
Range	7	30	40	43	43	40	2.76
Minimum	18	50	50	45	52	40	1.23
Maximum	25	80	90	88	95	80	3.99

Figure 1: Histogram for CGPA



The students are also encouraged to plot the histogram of CGPA, as shown in figure 1, using Excel. The diagram shows that 420 of the students are at-risk, which accounts for 21% of the overall population. There are also 502 students (25%) whose CGPA is between 3.5 and 4.

4. MODEL DEVELOPMENT & EVALUATION

In the next step, students are required to develop the predictive models for CGPA using any data analytics software. Students will develop the analytics models using the IBM-SPSS modeler for this course.

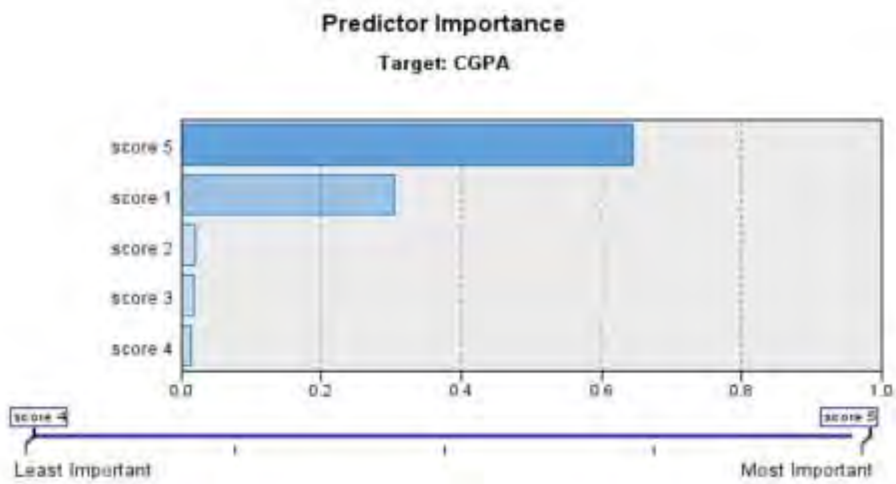
Next, we need to define the target variable as CGPA, a continuous value, and the rest as input variables. Next, change the Role of the Student ID to Record ID, gender to flag (0, 1), and score1 to score5 as continuous input variables. Finally, partition the data into 70% training and 30% testing.

Three predictive models will be used to predict CGPA. They are the decision tree method (DT), regression (Reg), and neural network (NN). The students will discuss the details of each model subsequently.

a. Model 1 - Decision Tree (CART)

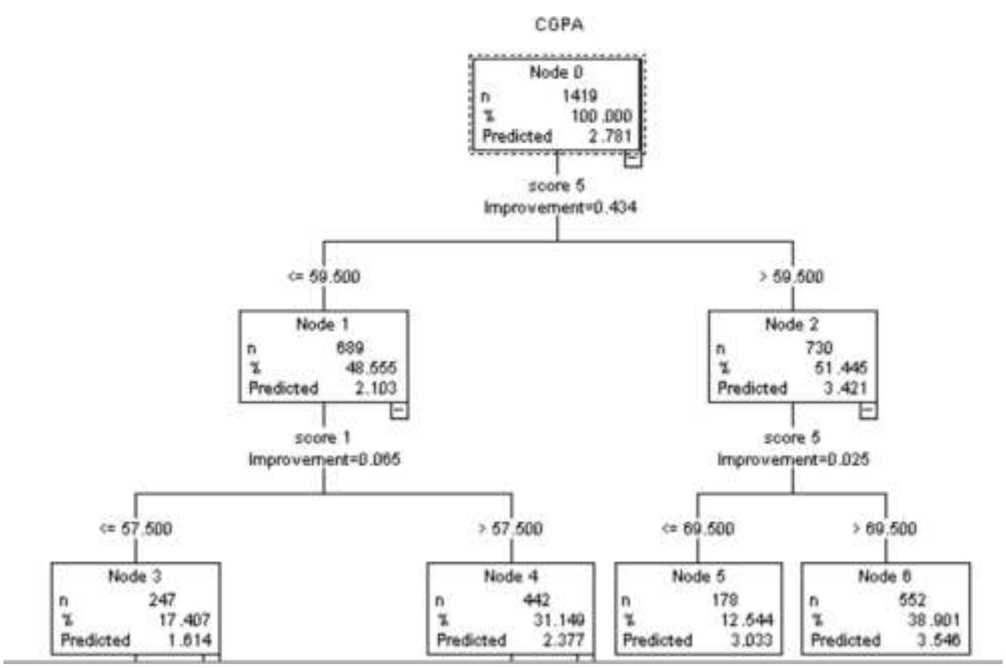
The three most important parameters given by the decision tree model are score5, score1, and score2.

Figure 2: Predictor importance for CGPA



The following chart is a partial tree that is generated using IBM-SPSS. The first variable that branches out is score 5, followed by score 1.

Figure 3: Partial Decision Tree diagram



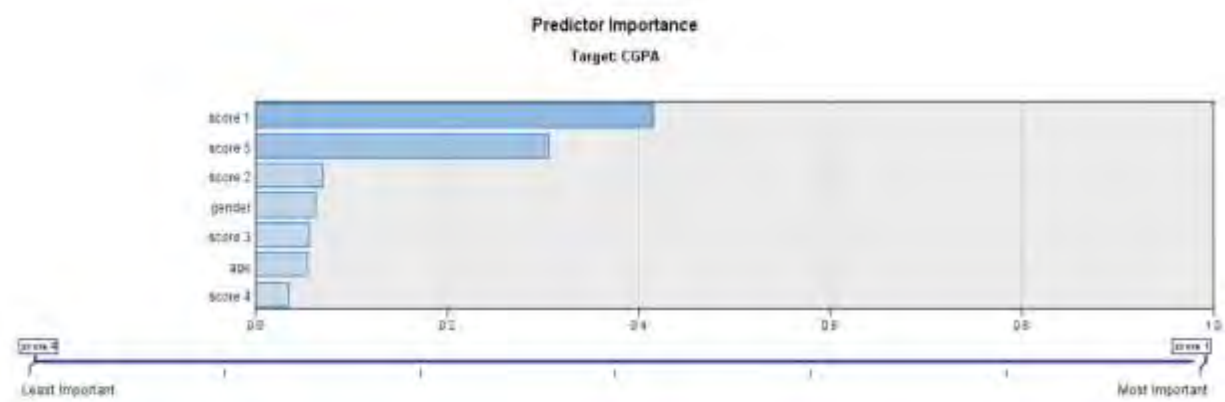
Some of the Decision rules are:

If score 5 > 59.5 and score 5 <= 69.5
CGPA = 3.033
Else if score 5 > 69.5
CGPA = 3.546

b. Model 2 – Regression (Reg)

The three most important parameters for the regression model, are score 1, score 5, and score 2.

Figure 4: Predictor importance for CGPA



Model output is as in Table 3 below.

Table 3: Regression model output				
Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.971[a]	.942	.942	.181467

ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	734.374	7	104.911	3185.857	.000[b]
	Residual	45.081	1369	.033		
	Total	779.455	1376			

a. Dependent Variable: CGPA

b. Predictors: (Constant), score 5, score 3, score 2, age, gender, score 4, score 1

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2.471	.080		-30.817	.000
	gender	.060	.010	.037	5.684	.000
	age	.002	.002	.005	.798	.425
	score 1	.045	.001	.558	31.202	.000
	score 2	.003	.000	.054	8.277	.000
	score 3	.003	.000	.054	8.299	.000
	score 4	.003	.000	.065	9.979	.000
	score 5	.024	.001	.432	24.174	.000

a. Dependent Variable: CGPA

The adjusted R^2 is 0.942, which shows that the regression model is a good fit.
The Sum of square error is 45.081, and the mean square error is 0.033.

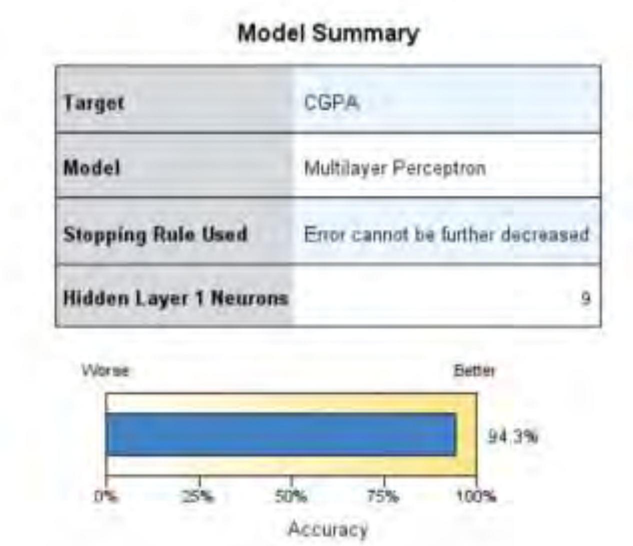
All the variables are significant in the regression model, and the coefficients are given in the table above. One unit increase in score 1, CGPA will increase by 0.045 and one unit increase in score 5 will also increase CGPA by 0.024. Thus, these are two most important input parameters for CGPA.

The equation of the regression is:

CGPA = - 2.471 + gender * 0.060 + age * 0.002 + score 1 * 0.045 + score 2 * 0.003 + score 3 * 0.003 + score 4 * 0.003 + score 5 * 0.024

c. Model 3 – Neural Network (NN)

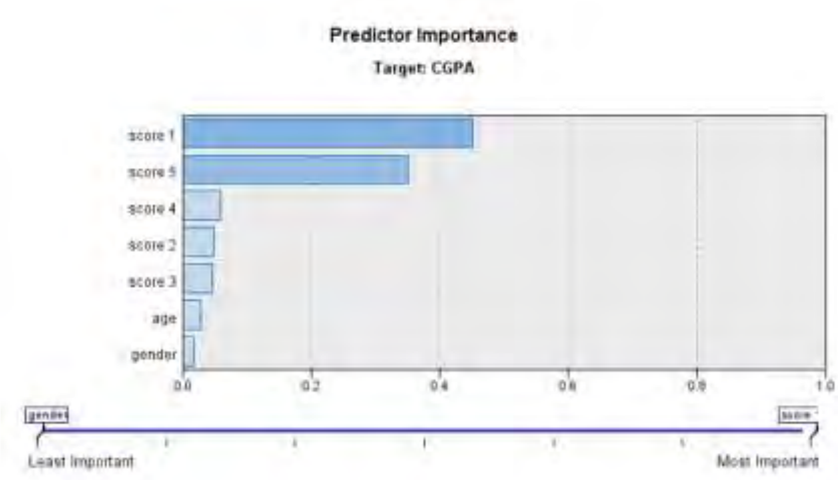
Figure 5: Model summary from Neutral Network



Model Summary

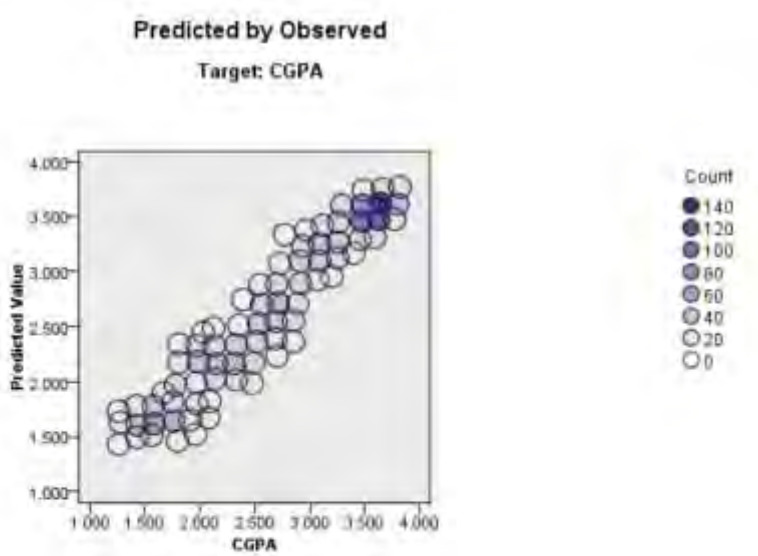
Target	CGPA
Model	Multilayer Perceptron
Stopping Rule Used	Error cannot be further decreased
Hidden Layer 1 Neurons	9

Figure 6: Predictor importance of Neural Network



The top three important variables from the neural network are score 1, score5, and score 4.

Figure 7: Predictive vs. actual CGPA from Neural Network



The model accuracy rate is 94.3%, which shows that NN is an excellent predictive model for CGPA.
After the students have built the predictive models, they need to evaluate the model performance and compare and determine the best model for implementation or use in the real world.
As the target variable CGPA is a continuous variable, they may use one of the performance indicators such as the Mean Absolute Error (MAE) from each model to compare. They will choose the model with the lowest MAE value as the champion model. Table 4 shows the models' comparison results.

Table 4: Models comparison

	Decision Tree (DT)	Regression (Reg)	Neural Network (NN)	Best Model
Minimum Error	-0.569	-0.831	-0.564	
Maximum Error	0.575	0.779	0.582	
Mean Error	-0.0	0.003	-0.004	
Mean Absolute Error (MAE)	**0.147**	**0.138**	**0.145**	**Regression**
Standard Deviation	0.182	0.178	0.181	
Linear Correlation	0.971	0.972	0.971	

Based on the MAE, the regression has the lowest value of 0.138 compared to neural network 0.145 and decision tree 0.147. The regression model assumes that the input variables such as age, gender, and score1 to score5 have a linear relationship with the target variable CGPA. The independent variables should not be correlated; the residues are independent and follow the normal probability distribution. The students also noted that regression is one of the most commonly used models in the real world as it is understandable and explainable by the users, the students have proposed to use the regression model in the real world.

5. INTERVENTION METHODS

The most widely used learning analytics application is the early identification of at-risk students. At-risk is a big issue in most universities around the world. University has past students' academic performance, which can be used to build the model to identify students at risk based on CGPA. Using the data analytics method that has been developed above to identify the at-risk students early, the management can take early intervention in reducing the number of at-risk students and the failure rate of students. As score5 and score1 are more critical input parameters with a high coefficient value that will affect the CGPA, we want the students to have a good score for these two subjects.

Similarly, we can also use a decision tree to come to the same conclusion. One of the rules is that, if the score5 is less than or equal to (<=) 59.5 and the score1 <= 57.5, the students' CGPA will be less than 2.0. Thus, faculty members teaching the course can monitor the students who did poorly in subject5 and 1. In addition, the instructors can set up additional consultation hours with students who need help to go through complex topics and improve their overall grades.

The head of programs can also use the information to notify the students if their CGPA is below 2.0 for the consecutive two terms. Set up a consultation session with the students to review their academic performance and direct them to the necessary peer support and online support. Based on the authors' experience, with the proper intervention methods at the appropriate time, the student's success rate has been improved by over 10% in the case study. This early intervention is undoubtedly critical as it can help students succeed academically and enhance the university's reputation.

Although predictive analytics can inform management on the possible at-risk students, they need to manage the potential emotional impact of labeling students at risk, traumatizing some students (Dziuban, Moskal, Cavanagh & Watts, 2012). Some students might feel demoralized and drop out. Therefore, instructors need to exercise care to build students' self-esteem while honest with their feedback. One suggestion is that the instructor can have a mentoring session with at-risk students to understand their challenges and encourage them.

Lastly, the analytics models developed are not 100% accurate; the predictive model may be wrong and may give a possible false error to state that predicted students are ok, but they are at risk. As such, it is also unethical to label students at risk prematurely. Therefore, institutions should always keep an open mind, exercise sound judgment when using this tool, and not be overly dependent on the predictions.

6. CONCLUSION

The case is taken from Education analytics for adult learning course, which is geared towards adult learning and self-directed learning. We use data analytical tools like Spreadsheet Excel and IBM SPSS Modeler to apply some data mining techniques to solve a business problem related to educational analytics. Students are given access to an e-textbook as one of the reading materials before the course and will need to spend 4-5 hours of preparatory work before the class. There will be in-class discussion and reinforcement of knowledge during the face-to-face lesson through lectures and demonstrations. Most importantly, we will apply some data mining techniques to derive some insights from the data to solve some business problems in educational analytics. Course participants are strongly encouraged to have an active learning and interactive classroom environment.

This course allows students to analyze educational data and make better decisions. This case study developed here serves as a tool to give students a real-world to use real-world data and analyze them in the control environment.

REFERENCES

Aldowah, H, Al-Samarraie, H. and Wan Mohamad, F, (2019). “Educational data mining and learning analytics for 21st century higher education: A review and synthesis”, Telematics and Informatics, **3**, 13-49.

Chiappe, A., & Rodríguez, L. P. (2017). “Learning Analytics in 21st century education: A review.” Ensaio: Avaliação e Políticas PúblicasemEducação, (AHEAD).

Dziuban, C., Moskal, P., Cavanagh, T., & Watts, A. (2012). “Analytics that Inform the University: Using Data You Already Have.”, Journal of Asynchronous Learning Networks, 16(3), 21-38.

H.S.Barrows (1996), “Problem based learning in medicine and beyond: A brief overview”, New Directions for Teaching and Learning, 68, 3-12.

Ma Nang Laik, Ivy Chia (2020), "Designing Learning Analytics for Effective Learning Through Problem-Based Learning", Journal of Education and Practice, Vol 11, No 18 (2020).

Reynol Junco, Candrianna Clem (2015), “Predicting course outcomes with digital textbook usage data”, The Internet and Higher Education, Vol 27, 54-63

Sclater, N. (2017). Learning Analytics Explained (1st ed.). Routledge.
https://doi.org/10.4324/9781315679563

DECISION SCIENCES INSTITUTE

Using vertical Federated Learning in industrial Supply Chains

Jonas Kallisch
University of Applied Sciences Emden/Leer
Email: jonas.kallisch@hs-emden-leer.de

Christoph Wunck
University of Applied Sciences Emden/Leer
OFFIS Institute for Information Technology Oldenburg
Iowa State University
Email: christoph.wunck@hs-emden-leer.de

ABSTRACT

This paper discusses the usage of Federated Learning, especially vertical federated Learning in industrial Supply Chains. The paper discusses why companies need a solution to securely analyze data together with partners in their Supply Chain. In the second part of the paper, we will look at the potentials of Federated Learning for analyzing intercompany data. The last part of the paper presents solutions to implement Federated Learning in an example industrial Supply Chain.

KEYWORDS: Supply-Chains, Federated Learning, Intercompany relations, Digital business ecosystems, Decision support systems, Data ownership

INTRODUCTION

The benefits offered by analysis algorithms have become increasingly relevant in the optimisation of operational processes and manufacturing activities within companies. (Lade, Ghosh, & Srinivasan, 2017) The results of the analyses can lower reject rates, reduce costs or even increase product quality. However, each company only applies these algorithms on its own structures and processes, which leads to local instead of a supply-chain wide optimisation of the production. (Kozlenkova, Hult, Lund, Mena, & Kekec, 2015)

In addition to technical difficulties and a lack of experts, many companies are concerned about the loss of control over their data, as a series of interviews conducted by the authors and corresponding sources prove. (Kallisch & Wunck, 2022) (Södergren & Cartling Wallén, 2022)

Approaches to exchange data between companies are not supported or asked by entrepreneurs, especially those of small and medium-sized enterprises. (Kallisch, & Wunck, 2022) The main reason for their concern, as numerous studies have shown, is the danger that third parties, including competing companies, could come into possession of the data and thus obtain sensitive company information. (Colicchia, Creazza, Noè, & Strozzi, 2019) This would endanger the competitive position of the companies. Even the transfer of data inside the supply-chain can pose risks for the companies sharing the data. (Panahifar, Byrne, Salam, & Heavey, 2018) For example, information about their cost structures and utilisation could be determined from knowledge about processes and thus be used by malicious partners to gain advantages. A loss of data by one of the partners would also be a problem. This could happen for example if the IT-security of one supply-chain member is insufficient. For these reasons, there is a need to protect the interests of a data provider, even within a trusted supply chain.

Several options are possible to meet this challenge. (Kallisch, & Wunck, 2022) In a previous paper, the authors have already identified and presented various existing approaches to solving the problem based on a literature review and expert interviews.

One of the identified options for a data sovereignty protecting cross-company analysis of data is the usage of federated learning. Federated learning offers the chance of analysing data that is not stored central. This feature serves to minimise the sharing of data between data stores, but also to enable devices to analyse data without training the algorithm used. The most use cases for federated learning are to enable devices such as edge devices or smartphones to analyse data without having to train the analysis algorithm themselves.

Federated learning also offers the advantage that data sets can be analysed without having to merge them first. This is crucial for companies, as it allows them to retain sovereignty over their data and the associated intellectual property. For many companies, this intellectual property represents the real value of the company. As a result, many companies are sceptical about sharing data, even within the supply chain, foregoing the potential added value that comes with the sharing of the data. In order to understand this problem more clearly and to answer the question of whether federated learning can contribute to solving the problem, we will first look at the extent to which data is already being shared between companies.

Current Supply Chain intercompany data analytics

The current state of intercompany data analytics in a supply chain can be considered as not regularly used. (Stock & Boyer, 2009) This means, that the data exchange between the companies in supply chain is often limited to the business communication, which means that the companies interact via orders, invoices or other logistical data elements. All this happens on the level of the Enterprise Resource Planning (ERP) system. The Manufacturing Execution System (MES) and the systems on the shop floor are usually not involved in the exchange.

Their task is to communicate with the suppliers and customers in vertical communication by sending material requirements or by converting customer orders into production orders and later into individual machine orders. The result is that neither the participating companies nor the supply chain management can access all data that might be available on a certain step inside the supply chain. (Y. Xu & Chen, 2018)

A communication between the different companies on the manufacturing level is not regulary used and therefore the analysis of the whole intercompany data along the supply chain. To better understand, why this is an issue and might even be a threat for some companies this paper will now describe the benefits of intercompany data analytics.

Benefits of intercompany data analytics

A key advantage of intercompany data analysis is that it can reveal connections between manufacturing activities in one company and the results of processing their products in another. (Rong, Hu, Lin, Shi, & Guo, 2015) Companies can thus identify trends, patterns and correlations that they would otherwise have missed. This enables better decision-making and strategic planning.

The combined analysis of data from different sources of the companies involved can also identify and fix data quality problems. (Madnick & Zhu, 2006) For example, if one company has erroneous data, the data from other companies can help identify and correct these errors. Therefore, the sharing of data between companies can even improve existing processes inside of a company that are not associated with the supply chain itself. In addition to these benefits, using a larger data set can improve the accuracy of data analysis. The collaborating partners can make more accurate predictions and more informed decisions. Analysis can also be used to identify new business opportunities and market trends.

As a last point, data analytics can also reduce companies costs, as it theoretically enables the sharing of resources and infrastructure. (Alcácer & Cruz-Machado, 2019) This can lead to cost savings, as companies no longer need to invest separately in expensive analytical tools and experts for data analysis.

Project “Zukunftslabor Produktion” Lower Saxony

To support the digitization of businesses the government of Lower Saxony in Germany has started the Project “Zukunftslabor Produktion” [Futurelab Manufacturing] (ZLP). (ZDIN, 2020) This interdisciplinary project is developing solutions and models especially for small and medium businesses in the manufacturing sector. Based on a use-case in the production of die-cast aluminium parts, the project examines how technical manufacturers can use data analytics and other industry 4.0 methods to improve their shop floor systems and create more resilient supply chain. The subproject “infrastructure and security” investigates concepts for connecting shop floor data of these independent businesses. A key focus lies on the security of the data, as safeguarding the intellectual property of the companies is a key asset of their business model. This paper will discuss the usage of Federated Learning inside of Supply Chains.

Research Questions and methods

In this paper, we try to answer the research questions:

1. Is federated learning a suitable solution to address the challenges of intercompany data analytics in supply chains?
2. Are there existing examples of federated learning in supply chains that can act as models for widespread use?
3. How can applications of federated learning be integrated in a supply chain and what value does it deliver to the companies?

A literature review was conducted to answer the first two research questions. In order to identify relevant papers the portals ScienceDirect and Google Scholar have been used. ScienceDirect has been choosen, because of its extensive research collection, the mostly fully accessible papers and the extensive literature base for manufacturing and logistics management. Google

Scholar adds the opportunity to get papers, books and journals from more publishers, since it lists a wide range of sources, including academic publishers, universities, and institutional repositories. This extensive coverage increases the likelihood of finding relevant research materials.

For the research the search terms “Federated Learning” and “Federated Machine Learning” in combination with the terms “supply chain”, “value chain”, “business ecosystem” and “transition”. This search was supplemented with “data analytics”, “privacy analytics” and a combination of the terms “manufacturing network”, “supply chain”, “intercompany” and “third party”. We linked the terms individually with an AND.

We included a contribution if it was a document-type article, conference paper, or book chapter, and if the contribution established a connection to the topic of federated learning in companies. The review identified a total of 76 sources. Contributions had been excluded, if it was not written in english or german, had no named author, consisted only of an abstract or had not been peer-reviewed, leaving 26 papers found to be relevant.

To answer the third research question, a case study with data from an industrial company was conducted as part of the ZLP project. For this purpose, a prototype for such an application has been developed. This prototype will be presented in this paper. But first we will start with the literature review on federated learning. The prototype was tested on a distributed set of data, provided by the partner, containing more then 6.000 elements. All parts have been quality signed and could therefore be analysed.

LITERATURE REVIEW

Horizontal and vertical Federated Learning

The basic concept of federated learning is that the training of models happens on distributed devices without actually having to transfer the data to a central system. (Q. Yang, Liu, Chen, & Tong, 2019) Therefore, the difference to traditional analytical models is that the whole system trains the analytical model instead of a central supercomputer. As described in the introduction the transfer of data can be an issue, if the data contains sensitive or user related information. Using federated learning allows to share only the aggregated model properties between the interacting devices without exchanging the data itself.

By that, the most important advantage of federated learning is that the concept technically protects the privacy of the users. (Z. Liu, Guo, Yang, Fan, Lam, & Zhao, 2022) Since the data is not transmitted to a central server, there is no risk that the data can be viewed by potentially malicious partners within the supply chain or even be leaked to competing companies. Due to the nature of the algorithms used, there is also no need for participants to share their data with third parties in order to train the model. Thus, federated learning, through this better control over the flow of data, is able to address companies' concerns about losing control through a technical measure. However, this comes with a few drawbacks. On the one hand, it is possible that a decentralised analysis of the data of cooperating companies does not deliver the same quality of results as an analysis of the same data in aggregated form using classic methods of big data analysis. (Kairouz, et. al. 2019) On the other hand, the use of federated learning models comes with the issue that all participating companies must build up skills in data analysis. (Zhan, Li, Guo, & Qu, 2021) This leads to increased costs compared to the joint analysis of intercompany data. Finally, it also means that individual participants may not be able to provide the same skills in terms of analysing the data. This can also have a negative impact on the outcome. To understand how federated learning methods work, it is important to note that there are two types of federated learning. These are horizontal and vertical federated learning.

Horizontal federated learning applies in cases where the participants have similar types of data and want to gain common insights from them without merging them. Yang et. al. describes it as follows "Horizontal federated learning, or sample-based federated learning, is introduced in the scenarios that data sets share the same feature space but different in samples." (Q. Yang et al., 2019) According to this definition, the type of data of each participant must be very similar for this type of federated learning. In concrete terms, this means that the data to be analysed in this way must have a very similar structure, but may well take on different forms. The aim of this type of federated learning is either to analyse remote data separately but in the same way, or to enable devices to analyse data without generating the necessary algorithm models on their own.

A typical example of an application of horizontal federated learning are applications on smartphones that would not be able to train the applied algorithms themselves, but can execute them. Their algorithms, for example for recognising voices, languages or faces, are generated by a high-performance infrastructure provided by the developer. (T. Li, Sahu, Talwalkar, & Smith, 2019) The generated algorithms can then be executed with significantly lower performance. On the one hand, the use of this methodology offers the advantage that the users' data remain safe from unauthorised access and, on the other hand, relieves the capacities of central analysis units.

In addition, it should be noted that due to limited network capacities for some applications, a transfer of data is not possible from a purely technical point of view and therefore procedures for decentralised analysis of these are indispensable. One example of such an application related to the amount of data is energy supply companies. (Singh, Masud, Hossain, Kaur, Muhammad, & Ghoneim, 2022) These collect consumption data in a decentralised manner and analyse it to develop forecasts about their energy needs. The data from the regional sub-organisations cannot be forwarded to central analysis databases here, or not in the time needed to react. Decentralised analysis is therefore necessary. Similar use cases can be found in various production systems, for example in the analysis of machine data.

Another application of horizontal federated learning can be found in medical diagnostics. (Sheller, Edwards, Reina, et. al., 2020) In this field, it is particularly important that this involves personal patient data. These can therefore often not be shared. Vertical federated learning can offer an option for such data, for example in hospitals and other medical institutions, to analyse it across institutions. However, this use is only possible if the data of the different institutions contain a similar structure and thus distributed training is possible.

In summary, horizontal federated learning can provide solutions for analyses where the data structures of the different partners are very similar, but merging them is not possible for one or more reasons. This is typically not the case in supply chains, where the different supply chain participants each have different data about their respective manufacturing processes. These data structures differ significantly because they are based on different systems, processes and procedures. Nevertheless, they are, typically, related to the same goal, usually an end product, and are therefore interrelated.

Vertical federated learning differs from horizontal federated as it changes the similar data structure through a similarity in the objects. Yang et. al. have described it as followed: "Vertical federated learning or feature-based federated learning is applicable to the cases that two data sets share the same sample ID space but differ in feature space." (Q. Yang et al., 2019) Therefore, the data structures of the data from the different participants differ much, but match in terms of the samples represented. It is important to note here that the participants must jointly identify the characteristics of the different datasets that are relevant for the analyses and the characteristics of these that match. This is necessary to identify the points at which the different participants can combine their data. Through a subsequent secure exchange of information, models can then be formed that allow for an overarching analysis of the data without actually transmitting the data.

In contrast to horizontal federated learning, an individual analysis of the individual data sets by the respective data owners is therefore necessary, which means that they must have the corresponding knowledge to adapt the respective models. (Khan, Thij, & Wilbik, 2023) However, depending on the approach, the degree of skill required may vary. First, the jointly developed model is trained on the own data sets, then the models of the individual participants are exchanged and transferred into a common analysis model.

As an example of the use of vertical federated learning, imagine the relationship between a bank and a credit card company. (Mammen, 2021) These could combine their data. The bank has information about account balances and transactions, while the credit card company has information about credit scores and payment histories. Through vertical federated learning, they can develop a joint model to better assess credit risk without looking at the specific personal data. Such an example shows that federated learning is able to support the cross-organisational analysis of sensitive and legally non-distributable data.

Another use case for vertical federated learning can also be the sharing of analytics between machines and/or workstations. (Niknam, Dhillon, & Reed, 2020) In this use case, the purpose of using vertical federated learning is predominantly not to preserve the data sovereignty of the participants, but rather to reduce the amount of data to be exchanged. Reasons for this can be, for example, the available bandwidth between the different machines or also production sites or also that the process time is not sufficient for a central analysis. This use case is thus similar to horizontal federated learning with the difference that in this case the connected devices are not similar and do not have a uniform data structure.

In summary, vertical federated learning offers the possibility to analyse data from different processes, machines or companies without having to make this data accessible. This makes this type of federated learning particularly suitable for use in cross-company data analysis. The first of the research questions mentioned can thus be confirmed.

However, research questions still remain as to how federated learning can be used in supply chains and what potential the use of federated learning models can offer. To do this, the relationship must first be clearly worked out and the impact of the structure of a supply chain must also be considered. To this end, we will first look at how federated learning models can be used in supply chain models for secure data exchange. Afterwards, a prototype developed within the framework of the Zukunftslabor Produktion project will be presented.

State of usage of Federated Learning inside of Supply Chains

After defining what is meant by federated learning in the previous section and confirming that the concept is capable of being used in cross-organisational data analysis, the question arises whether federated learning is already being used in practice. To answer this question, a second literature review was conducted. In this, the search terms "federated learning", "vertical federated learning" and "federated analytics" were combined with the search terms "supply chain", "value chain", "business ecosystems" and "business model". The identified literature was then analysed with a focus on examples of practical applications of federated learning. The result of the analysis showed that only a few cases of the use of federated learning in a cross-company context are publicly known.

One identified example of the use of federated learning in supply chains is in the agri-food sector. (Durrant, Markovic, Matthews, May, Enright, & Leontidis, 2022) This shows how federated learning can solve a problem of data availability. In this sector, the challenge is that different actors individually do not have sufficient amounts of data to benefit from machine learning. However, together they can create a common model and benefit from the common data set. The authors have empirically shown that the use of federated learning can produce improvements in this environment.

Another example can be found in the field of e-commerce. (J. Li, Cui, Yang, Yuan, He, & Li, 2021) Here, existing work shows that horizontal federated learning specifically in this area is able to significantly increase the accuracy of demand forecasting without sharing data between companies. While the study is based on only a few participants and is also largely based on public data, this nevertheless shows the potential benefits of federated learning.

In addition to these examples, some work exists on the use of federated learning models in the context of blockchains in different areas. (D. Li, Luo, & Cao, 2022) Examples include supply chain management, collaborative product design and Internet of Things systems.

In summary, there are several examples of the use of federated learning models in supply chains. However, the use of vertical federated learning is essentially treated theoretically. In some cases, companies, especially in the financial sector, are starting to build customised solutions to specific problems due to data sharing constraints. Cross-industry applications for federated learning in supply chains, on the other hand, are not available to companies. The question is how an application of federated learning could be brought in and how it can be used in supply chains.

How to use Federated Learning inside of Supply Chains

As already described, vertical federated learning is particularly suitable for use within a supply chain, as the companies in a supply chain do not have matching data structures. In order to carry out the analysis within a supply chain, the structures of the supply chain must first be recorded.

In principle, the supply chain, or the area of the supply chain intended for intercompany analysis, must first be defined. (Q. Yang et al., 2019) Within this group, the common target characteristics, the labels of the data, must then be determined and the relevant characteristics of the respective participants selected. Depending on the level of trust, the exchange of concrete data can increase the quality of the analysis. In addition, insight into the data can reduce the cost of the analysis, as the analysis models can be created jointly.

The findings of the ZLP so far show that it is easiest to link company data stocks via the physical flows of goods. (Kallisch, & Wunck, 2022) In these links there are usually already shared data points, such as serial or batch numbers. These can be clearly assigned between the companies and can thus be used for most federated analyses. Special cases, such as matching characteristics, must be considered separately.

In theory, networking and intercompany data analysis are therefore not a challenge. In practice, however, especially in small and medium-sized enterprises, there is a lack of specific technical knowledge and infrastructure to handle the communication, model aggregation and security measures that the use of federated learning requires. (Alazab, RM, M, Maddikunta, Gadekallu, & Pham, 2022) In addition to the know-how to analyse distributed data, the use of federated learning also requires the availability of data relevant to the supply chain. This means that cross-company analysis of data only makes sense if the companies involved have data. If one or more companies in the supply chain do not have access to such data, implementing federated learning can be difficult.

Another problem is the availability of products and experience with the results of using federated learning in supply chains. (Posner, Tseng, Aloqaily, & Jararweh, 2021) Some companies have developed solutions based on federated learning on an individual basis for individual use cases. Cross-cutting solutions for decentralised analysis of data between companies, on the other hand, are rare and only sector-specific. In contrast, cross-company platforms such as the International Data Spaces, GAIA-X or Catena-X are based on the concept of direct data exchange. From the authors' point of view, there is therefore a need for a solution for the use of vertical federated learning in industrial supply chains. We will present a prototype for such a cross-company platform for analysis in the following section.

PROTOTYP AND CASE STUDY

Prototype for a secured information exchange with vertical federated learning

To enable the use of federated learning models in supply chains, it is necessary to digital map these supply chains. On the one hand, this mapping must integrate the data of the individual partners and enable those responsible to configure them, and on the other hand, it must contain the relationships between the companies. These relationships between the companies and their data sources are the basis for configuring concrete models of cross-company analysis.

In the realisation of the prototype for the use of vertical federated learning, a configuration menu was developed for connecting new data repositories to a data aggregation platform provided by the prototype. This menu allows organisations to transfer data stores within their IT infrastructure to the platform. The relationships between these data are realised within the system in the form of a workflow, so the application requires knowledge about the process and about the data assets. This is a limitation because companies need to bring in sufficient expertise. The view of the set-up screen for this is shown in Figure 1 on the next page. We apologise that the interface language is German, as the application was developed for German SMEs. Data sources for the application can be added in the interface.

Figure 1: Screen to configure data sources

To define the relationships between supply chain participants, each participant defines its outgoing information and edits how incoming information is represented in its own data structure. Ideally, incoming information from suppliers corresponds exactly to the information stored in their own data structure. For example, the serial number of a component is also stored in its own structure and thus enables the direct assignment of all selected data points. Figure 2 shows the interface for configuring the supply chain. We apologise again that the interface is in German. This can be used to configure the entire supply relationship and filter the view by product or by relationship to a specific company. In this way, the supply chain as a whole can be viewed strategically and the decision of a data exchange can be considered with all interconnections in mind. Depending on the trust relationship, the platform can be configured to always show only the next and each previous stage of the chain.

Figure 2: Interface to configure flow of information

The concept of the application consists out of three layers:

- Communication layer
- Service layer
- Analytics layer

The first of these layers, the communication layer, is intended to establish data exchange between the individual enterprise data platforms. The concept for this is that the exchange of data within this layer is strictly oriented to the physical supply chain relationship, i.e., the exchange of goods and services. The communication layer is working is therefore configured with the represented interface in Figure 1 and Figure 2.

The second layer is used to implement services between the participating companies. These services allow companies to create individual applications for their supply chain. The layer in the interconnection platform describes these services and stores the rights of the respective participants to use the services. Requests to use services are forwarded via the communication layer and lead to the creation of an agreement to use the service. It would also be conceivable to extend the service system to an open system in which providers of data or analytics services could also interact in the supply chain. This layer is mainly responsible for enabling the federated learning, since the companies had to agree on sharing the information and define the level of information exchange with each other.

The final layer of the concept is for cross-enterprise data analytics. The layer is intended to provide a set of basic analytics, based on the concept of federated learning, which can be used by companies within their data platforms. The challenge in this application layer is to make the presentation and operation of the algorithms usable for a user who has no training in data analysis. To this end, we will conduct various studies in the future to be able to identify an interface that is as easy to understand as possible. No specific interface was created for the case study, as the configuration of the algorithms was not done by these users. In the following section of this paper we will describe the use case.

Case Study

The application of federated learning was tested in a case study. In this case study, data from a partner of the Future Lab Production was used. These come from two industrial processes. The first of these processes is the preparation and production of manufacturing tools. The second process uses the data from the first process to produce a product that is accepted by a customer. The linkage for federated learning can therefore be done by identifying the tool.

The data of the tool production contain structural information, such as the measured deviations from the design data, the life cycle of the respective tool or also the simulation data with parameters belonging to the mould. The data of the production contain structural data of the used equipment, the set parameters at the time of the respective production series, as well as time series data of the used sensor technology. In addition to this information, quality analysis data is available for the components.

In regular production, visual inspections of the manufactured products are carried out and, if deviations are detected, the parameters of the equipment are changed. If this test shows no effect, the production series is stopped and the tool is handed over to the tool designer for inspection in order to eliminate any damage or to recondition the tool. Thus, no analysis of the error sources is carried out.

To train the model, data points from approximately 8000 production cycles were used from production. A validation was carried out with the help of 2000 further data sets. A total of 58 measured variables were included in the analysis, distributed over the two data sources. The objective of the model is to determine the probability of errors occurring in production and whether it is possible to eliminate these errors by changing the system settings or whether it is necessary to repair the tool.

The analysis of the data was carried out with the help of the FedML framework, short for Federated Machine Learning, taking into account the separation of data availability. FedML has found applications in various domains, including healthcare, finance, and Internet of Things (IoT), where data is decentralized, sensitive, or subject to privacy regulations. It offers a promising approach to leverage the collective power of distributed devices while ensuring data privacy and ownership.

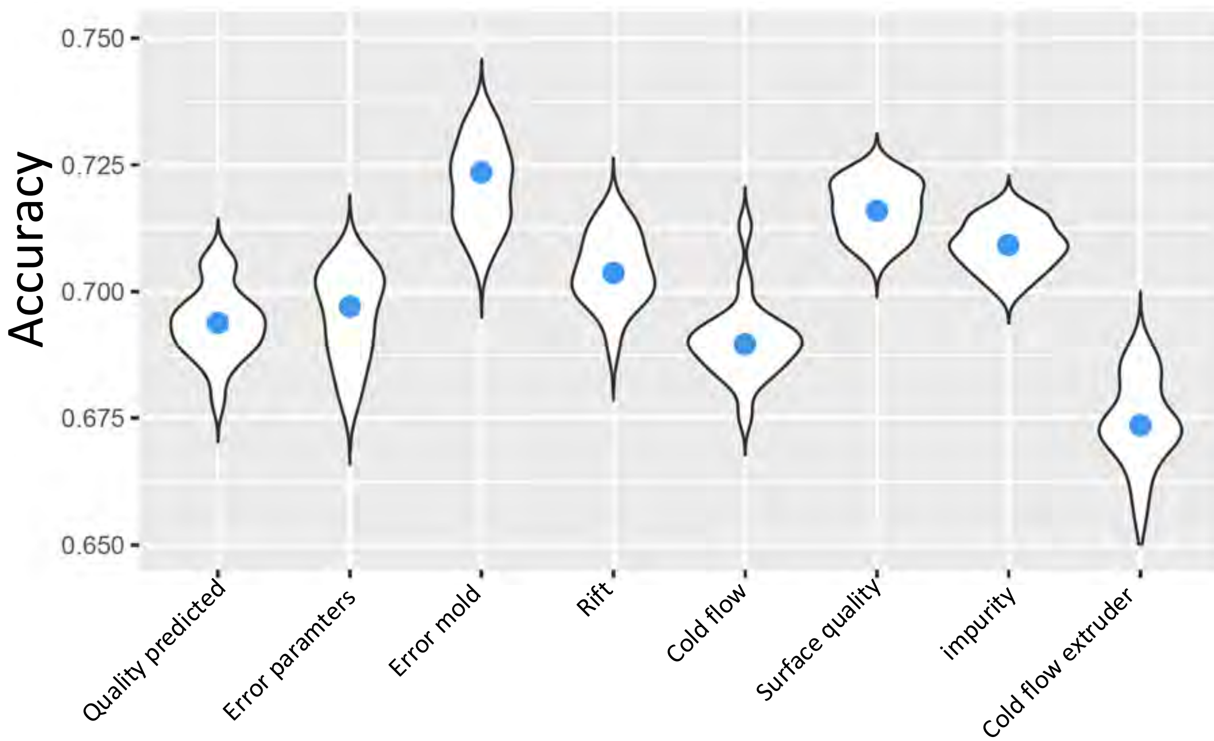


Figure 3 Results in case study

The accuracy of the generated model is shown in Figure 3. As can be seen, the quality of the prediction ranges between 65 and 75%. The prediction accuracy with regard to the statement whether a product meets the quality requirement is 69% on average, whereby errors in the mould shape are detected with a higher probability than those that lie in the parameters of the machine. The basis for the comparison is provided by the real assigned defects and defect parameters.

In summary, the analysis thus provides support for error detection, but it must be noted that the accuracy of the classification is not particularly high in the application case examined. Thus, federated learning can only serve as a supporting method in the use case. The decision-making process can thus be supported, but not taken over by the analysis algorithms. Nevertheless, the application offers the potential to decide more quickly whether a change in parameters is likely to be sufficient in the event of a fault or whether production must be stopped and the mould repaired. It is also conceivable that other data sources will be able to increase the prediction accuracy in the future.

Conclusion

In this paper, the use of vertical federated learning in supply chains was considered. For this purpose, the potentials and already existing as well as conceivable scenarios were determined within the framework of a literature analysis. The motivation for the use of federated learning, the protection of the intellectual property of the supply chain participants, is described and the causes of this are considered.

Existing applications of federated learning are discussed in order to answer the questions which applications can serve as a model for the implementation of federated learning in supply chains. It was found that the use of federated learning is usually limited to a few use cases and that mainly horizontal federated learning is used in industry. The paper shows how federated learning can be used in a cross-company context.

Finally, the paper presents a prototype application for the use of federated learning in supply chains. The prototype was tested in a case study and the results are presented. However, it must be noted that the results of the application in the supply chain described may not be transferable to other areas. There are actually several limitations in the case study. Firstly, the case study only looks at a section of a more complex supply chain. In a larger section or even in an entire supply chain, the results could differ significantly from those of the case study. Secondly, the case study uses data from very many sensors and other data sources. It is likely that not all supply chains have this high number of data points available from all participants. Lastly, the study does not represent the application of federated learning by the companies themselves. Thus, the analyses were not conducted without knowledge of the analytical models used. Therefore, further studies will be necessary in the near future to remove the limitations of this study and to gain further insights.

REFERENCES

Alazab, M., RM, S. P., M, P., Maddikunta, P. K. R., Gadekallu, T. R., & Pham, Q.-V. (2022). Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions. *IEEE Transactions on Industrial Informatics*, *18*(5), 3501–3509.

Alcácer, V., & Cruz-Machado, V. (2019). Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science and Technology, an International Journal*, *22*(3), 899–919.

Colicchia, C., Creazza, A., Noè, C., & Strozzi, F. (2019). Information sharing in supply chains: a review of risks and opportunities using the systematic literature network analysis (SLNA). *Supply Chain Management: An International Journal*, *24*(1), 5–21.

Durrant, A., Markovic, M., Matthews, D., May, D., Enright, J., & Leontidis, G. (2022). The role of cross-silo federated learning in facilitating data sharing in the agri-food sector. *Computers and Electronics in Agriculture*, *193*, 106648.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., & Bhagoji, A. N. et al. (2019). Advances and Open Problems in Federated Learning.

Kallisch, J., & Wunck, C. (2022). Options for connecting decentralized data infrastructure to improve Supply-Chain decision making without giving up individual data property.

Khan, A., Thij, M. ten, & Wilbik, A. (2023). *Vertical Federated Learning: A Structured Literature Review.*

Kozlenkova, I. V., Hult, G. T. M., Lund, D. J., Mena, J. A., & Kekec, P. (2015). The Role of Marketing Channels in Supply Chain Management. *Journal of Retailing*, *91*(4), 586–609.

Lade, P., Ghosh, R., & Srinivasan, S. (2017). Manufacturing Analytics and Industrial Internet of Things. *IEEE Intelligent Systems*, *32*(3), 74–79.

Li, D., Luo, Z., & Cao, B. (2022). Blockchain-based federated learning methodologies in smart environments. *Cluster Computing*, *25*(4), 2585–2599.

Li, J., Cui, T., Yang, K., Yuan, R., He, L., & Li, M. (2021). Demand Forecasting of E-Commerce Enterprises Based on Horizontal Federated Learning from the Perspective of Sustainable Development. *Sustainability*, *13*(23), 13050.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2019). Federated Learning: Challenges, Methods, and Future Directions. *3.*

Liu, Z., Guo, J., Yang, W., Fan, J., Lam, K.-Y., & Zhao, J. (2022). Privacy-Preserving Aggregation in Federated Learning: A Survey.

Madnick, S., & Zhu, H. (2006). Improving data quality through effective use of data semantics. *Data & Knowledge Engineering*, *59*(2), 460–475.

Mammen, P. M. (2021). *Federated Learning: Opportunities and Challenges.*

Niknam, S., Dhillon, H. S., & Reed, J. H. (2020). Federated Learning for Wireless Communications: Motivation, Opportunities, and Challenges. *IEEE Communications Magazine*, *58*(6), 46–51.

Panahifar, F., Byrne, P. J., Salam, M. A., & Heavey, C. (2018). Supply chain collaboration and firm's performance. *Journal of Enterprise Information Management*, *31*(3), 358–379.

Posner, J., Tseng, L., Aloqaily, M., & Jararweh, Y. (2021). Federated Learning in Vehicular Networks: Opportunities and Solutions. *IEEE Network*, *35*(2), 152–159.

Rong, K., Hu, G., Lin, Y., Shi, Y., & Guo, L. (2015). Understanding business ecosystem using a 6C framework in Internet-of-Things-based sectors. *International Journal of Production Economics*, *159*, 41–55.

Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., & Kotrotsou, A. et al. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, *10*(1), 12598.

Singh, P., Masud, M., Hossain, M. S., Kaur, A., Muhammad, G., & Ghoneim, A. (2022). Privacy-Preserving Serverless Computing Using Federated Learning for Smart Grids. *IEEE Transactions on Industrial Informatics*, *18*(11), 7843–7852.

Södergren, F., & Cartling Wallén, M. (2022). CREATING VALUE THROUGH INFORMATION SHARING : Exploring the Transition Towards a Digital Supply Chain. Retrieved from https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1654203&dswid=7113

Stock, J. R., & Boyer, S. L. (2009). Developing a consensus definition of supply chain management: a qualitative study. *International Journal of Physical Distribution & Logistics Management*, *39*(8), 690–711.

Xu, Y., & Chen, M. (2018). An Internet of Things based framework to enhance just-in-time manufacturing. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *232*(13), 2353–2363.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST) Volume 10 Issue 2, Article No. 12, January*, (12).

ZDIN (2020). PRODUKTION - ZDIN, accessed May 31, 2023, available at https://zdin.de/zukunftslabore/produktion.

Zhan, Y., Li, P., Guo, S., & Qu, Z. (2021). Incentive Mechanism Design for Federated Learning: Challenges and Opportunities. *IEEE Network*, *35*(4), 310–317.

DECISION SCIENCES INSTITUTE

Validation for Risk Severity Planning System with Multicollinear and Autoregressive Inputs

John Maleyeff
Boston University Metropolitan College
Email: maleyeff@bu.edu

Tianyu Xu
Boston University Metropolitan College
Email: tianyuxu@bu.edu

Tianyuan Liu
Boston University Metropolitan College
Email: lty2020@bu.edu

Jingyu Wang
Boston University Metropolitan College
Email: jingyuw@bu.edu

ABSTRACT

This article presents a methodology for modeling hourly weather variables that will be used to validate a severity determination risk prediction model that was developed for the MBTA in Boston. The approach uses a simulation that requires stochastic modeling of five climatic input variables that exhibit both multicollinearity and autocorrelation. Historical data are analyzed, and a regression is used to model daily temperature, humidity, windspeed, and precipitation. Clustering is used to classify hourly patterns for each weather variable with some sets of clusters exhibiting dependencies. The simulation was validated using face validation with MBTA experts.

KEYWORDS: Risk analysis, Simulation, Validation, Multivariant input modeling, Clustering, Public transit

INTRODUCTION

The public transit system in greater Boston is operated by the Massachusetts Bay Transportation Authority (MBTA). The MBTA is the sixth-largest system of its kind in the U.S., with 5 subway lines, 170 bus routes, 13 commuter rail lines, and 3 ferry routes. Its vehicles operate on most streets in Boston, with over 700 miles of subway and commuter rail tracks. MBTA facilities include storage and repair garages, over 6000 bus stops,140 commuter rail stations, and 148 subway stations (some of which are above ground). Finally, there are scores of MBTA parking lots that can hold almost 45,000 vehicles. Although the City of Boston is responsible for removing snow from streets and treating roads, the MBTA needs to mobilize its own resources and make schedule modifications in the event of a winter storm.

Boston is located near the Atlantic Ocean, where winters range from mild (e.g., total snowfall was only 11.4 inches in 2023) to harsh (e.g., total snowfall was 108.6 inches in 2015).

Temperatures can vary even during a single day, which can create icy roads when temperatures drop after a rainstorm. Snow can be heavy making it difficult for buses to move and winds can be extremely strong, especially for storms that originate from the northern Atlantic ocean. Boston is dominated by narrow streets that tend to change directions with short blocks and numerous irregular intersections with a dense, commercial core. Although this layout provides a quaint setting for tourists, it can be challenging for a public transportation entity.

Experienced MBTA managers tend to have a good understanding of how weather conditions will affect MBTA operations. As a storm approaches, its overall severity is assessed based on National Weather Service (NWS) forecasts, and individual managers take action to modify at-risk bus routes, notify snow removal contractors, mobilize its bus stop and subway station treatment personnel, and allocate its own snow removal personnel, among many other actions. Although the MBTA winter storm planning system is well established, some features require improvement. For example, its standard documentation includes a formalized system that focuses on snowstorms only, with individual managers taking a more nuanced approach for other upcoming weather conditions such as freezing rain, heavy winds, and snow fluffiness (where both very heavy snow and very light snow can cause difficulties to their operations).

The authors have been working with the MBTA risk and emergency managers to create an analytically sophisticated system to support MBTA decision makers. Its severity predictions concern a wide range of specific risks, and they are visualized on an hourly basis to enhance the timing of decisions. Preliminary testing of the models took place in the winter of 2021-2022 and a prototype system was implemented in Tableau software. Plans were made to test the prototype system more comprehensively with MBTA operations managers in the winter of 2022-2023. Unfortunately, the winter proved extremely mild with little snow and mild temperatures.

The purpose of this article is to show how the MBTA winter storm planning prototype system was simulated and validated based on a realistic range of Boston's winter weather scenarios. Two challenges existed. First, the models make use of daily weather variables that exhibit multicollinearity and hourly patterns that have an autoregressive relationship. Second, model predictions cannot be compared to "correct" values. This article includes a literature review placing the work in an academic context and highlighting the gaps in prior research that needed to be addressed, Then, a comprehensive set of Boston historical winter weather data was analyzed to determine how the input models for the simulation needed to be developed, using distribution fitting, regression techniques, and clustering. The models are presented, and the preliminary tests of the system are described. The article concludes with future research directions.

LITERATURE REVIEW

Modeling weather conditions is complex because the various parameters are not statistically independent, and their relationships can be location-dependent (Steadman, 1984). Therefore, any analysis needs to begin by describing the statistical properties among weather parameters (Wilkes & Willy, 1999). An early work proposed a Markov chain-exponential model to describe daily precipitation independently, then generate temperature and solar radiation (Richardson, 1981). Killsby et al (2006) proposed a weather generator model based on a regression based on daily climatic variables and daily rainfall. Previous works largely focused on daily, monthly, and annual weather generation for the study of climate impact (Srikanthan & McMahon, 2001).

More recent work finds that a strong correlation exists between hourly temperature and hourly relative humidity (Guan et al, 2007). Zeng (1999) also discovered a relationship between precipitation and temperature. Back & Bretherton (2005) noticed that a correlation between wind speed and precipitation was more pronounced at higher relative humidities. Finally, Bretherton (2004) described relationships between humidity and precipitation over ocean regions, which may be the case in Boston.

Simulating the winter storm planning system's logic requires the modeling of a multivariate input process based on historical data that consists of multiple dimensions or features (Biller and Ghosh, 2006). Biller (2009) developed a robust framework for multivariate time-series inputs that aims to model complex interactions and dependencies called Vector-Autoregressive-To-Anything, or VARTA. Bayesian methods have also been used for testing new observations in a multivariate input/output process for a fault control system (Gorinevsky, 2015). Barančoková & Kenderessy (2014) provide an example of the application of multivariate input modeling for simulation of a landslide susceptibility prediction simulation. A multivariate sensitivity analysis technique for dynamic models with random process inputs was presented by Li et al (2020).

K-means cluster analysis is a commonly employed technique to classify groups of entities that possess similar features (Sokol, 1974). The method employed here includes normalization and preprocessing of features, which is often used in clustering (Ren & Barnett, 2022). Visualizations are used to help an analyst determine the best number of clusters based on group homogeneity that favors more clusters and practical considerations that favor fewer clusters (Kumpf et al, 2018). Metsalu and Vilo (2015) developed a web tool to help with this determination using a heat-map and principal component analysis. For more complex applications, k-means clustering requires that an initial cluster assignment be carefully chosen (Bradley & Fayyad, 1998). Clustering analysis and regression models have been used to rectify null values in weather data sets (Sanhudo et al, 2020).

The aim of validation is to verify that the model aligns with the underlying logic and understanding of the real-world system. Validation is a crucial step in any data analysis process, particularly in the context of machine learning and statistical modeling, by ensuring that results produced by a model are accurate, reliable, and generalizable to other data sets (Naylor, 1967). There are various validation techniques, and they are recognized and utilized based on operational validity and data validity. When there is no precise known solution, face validation can be employed whereby individuals knowledgeable about the system are asked whether the model and/or its behavior are reasonable (Sargent, 2010). It has been used to validate a model that predicted a quality-of-life metric (Carlton, 2022), and to validate a robotic surgical simulator (Seixas-Mikelus, 2010).

METHODOLOGY

Development of the models used within a winter planning risk severity decision support system has been reported previously (reference held to protect anonymity). The system used machine learning methods to predict hourly severities associated with snow weight (affects manpower needed to remove snow), snow fluffiness (can clog train engines air intake valves), road slickness, parking spaces lost to snow piling, flat ice buildup (especially on in-ground transmitters), and cylindrical ice quantities (on trees and power lines). Several of these severities are based on a model that predicts the snow-to-water ratio (SWR). These severity predictions utilized machine learning methods that used historical NWS forecasts as predictor variables, and numerous response variables including road conditions found in the City's car crash database, road friction data from the Massachusetts Department of Transportation, National Oceanic and Atmospheric Administration (NOAA) data, and relevant models from published studies. The five input parameters used in the winter storm planning system were hourly values for: (a) temperature, (b) wind speed, (c) relative humidity, (d) rain amount, and (e) snowfall amount.

The methodology used to test the severity predictions required the determination of input models to accurately generate hourly values for the five weather parameters. The weather data were collected from the NWS (1/1/2016 through 3/23/2023) for the greater Boston area during the winter months (December through March). There were a total of 931 days in the data set, each with 24 hours of historical weather for the five parameters.

The methodology included a two-part procedure as shown in Figure 1. The first part used a correlation analysis to determine the degree of multicollinearity among the daily weather variables. The hourly weather data were aggregated by day, and included the average temperature, humidity, and wind speed, and the total amount of rainfall and snow. The variables deemed independent were analyzed to determine their daily distributions. The other parameters were analyzed using a regression approach that predicted their values (and a level of variation) based on the parameters on which their model was based.

The second part of the analysis considered the hourly distributions of each weather parameter, which follow a time series relationship. The hourly data were standardized as follows. For temperature, humidity, and wind speed, each hour's deviation from the daily average was calculated, while for the amount of rain or snow, the percentage of each day's total corresponding to each hour was calculated. The result consisted of 931 data sets of hourly deviations or percentages with each data set consisting of 24 time series outcomes per day. Clustering was then applied to these data to determine the cluster identities for each of the five weather parameters.

RESULTS

Results are shown below that detail the development of the random model for the simulation input variables, with each of the two parts of the analysis documented separately. Then, their joint use in generating random inputs is described.

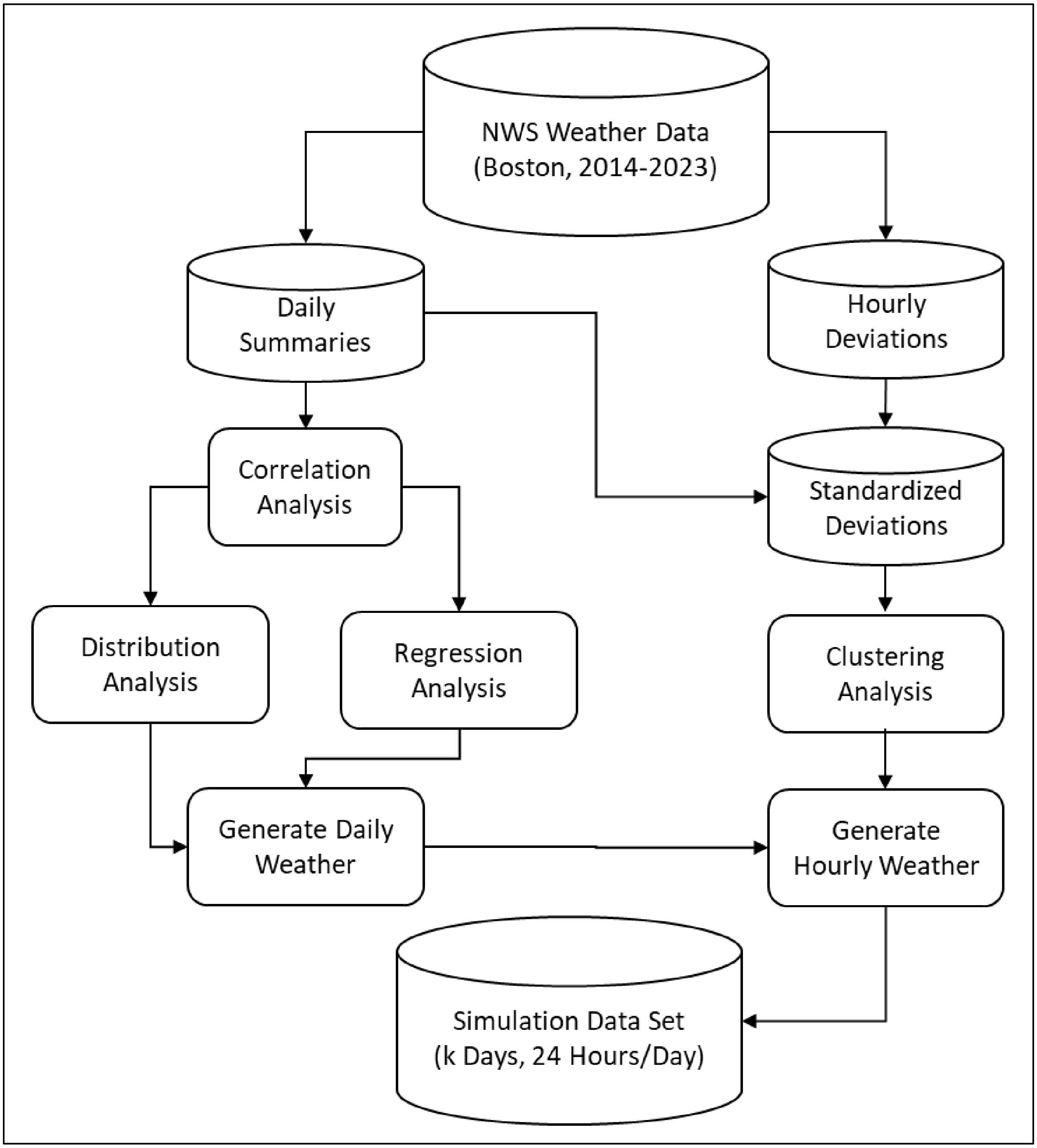


Figure 1: Methodology

Part 1 – Daily Predictions

Prior work suggests that the amount of multicollinearity needs to be considered among the daily weather parameters. Figure 2 shows the results of the correlation analysis (where the suffix "_A" represents the daily average of that parameter and "Pre_T" represents the daily total of precipitation). For each day, the precipitation was calculated as the total rain plus the estimated water content of the snow (using an average SWR of 10). Each variable has 931 total observations, so small correlations can be statistically significant but not especially meaningful (i.e., the correlation may be too small to influence the input parameters generation model). The results show that the pairs of parameters that exhibit significant and potentially important dependencies are: (a) precipitation and wind speed, (b) precipitation and temperature.

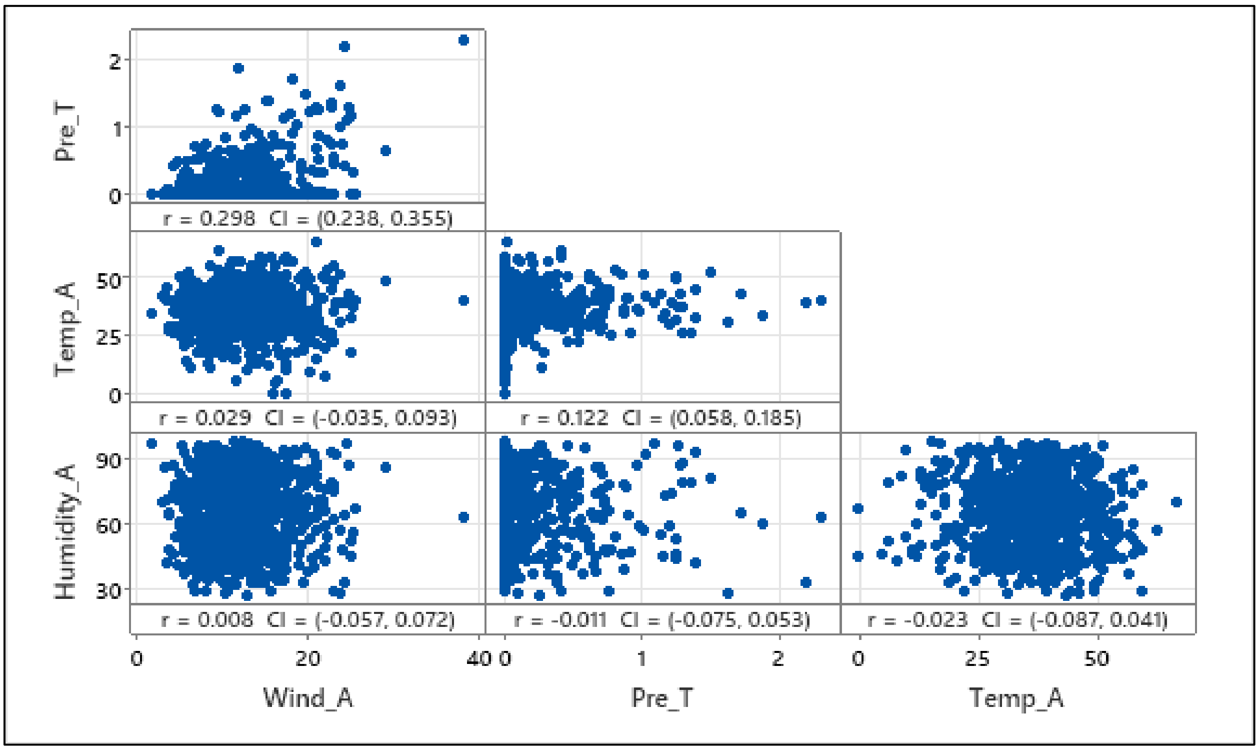


Figure 2: Correlation Results (r = correlation coefficient, CI = 95% confidence interval for r)

It appears that precipitation is dependent on temperature and wind speed, although the literature also suggests that precipitation can be correlated with these parameters as well as humidity (e.g., Zeng, 1999). Therefore, in the simulation's input model, the daily average values for temperature, wind speed, and humidity will be generated independently, while the parameter for daily total precipitation will be determined using regression analysis. Distribution analyses concluded that the three-parameter Weibull distribution was a suitable fit for the average daily winter temperature (Figure 3).

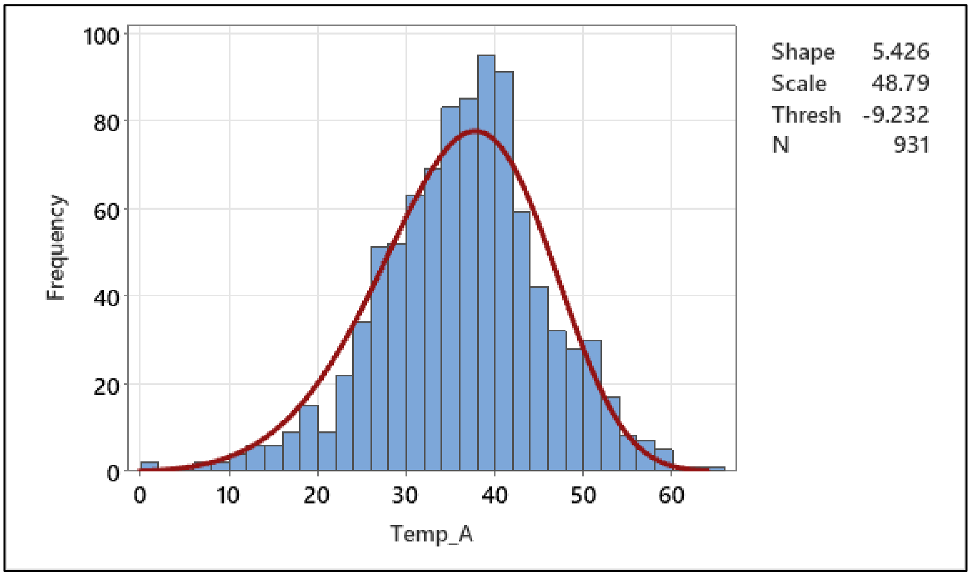


Figure 3: Distribution of The Average Daily Winter Temperature with Weibull Fit

In addition, the gamma distribution was a suitable fit for the average daily winter wind speed (Figure 4).

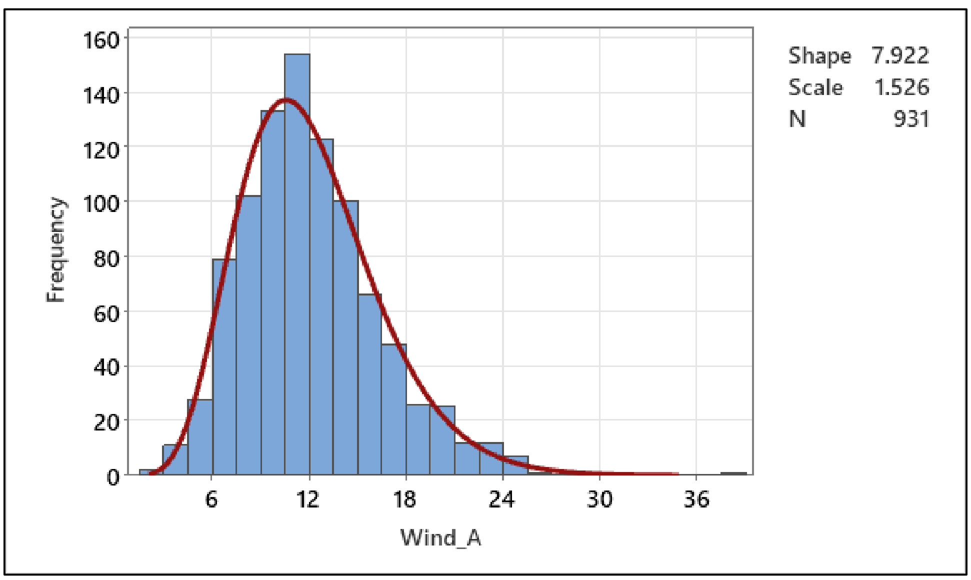


Figure 4: Distribution of The Average Daily Winter Wind Speed with Gamma Fit

And, a truncated normal was a suitable fit for the average daily humidity (Figure 5).

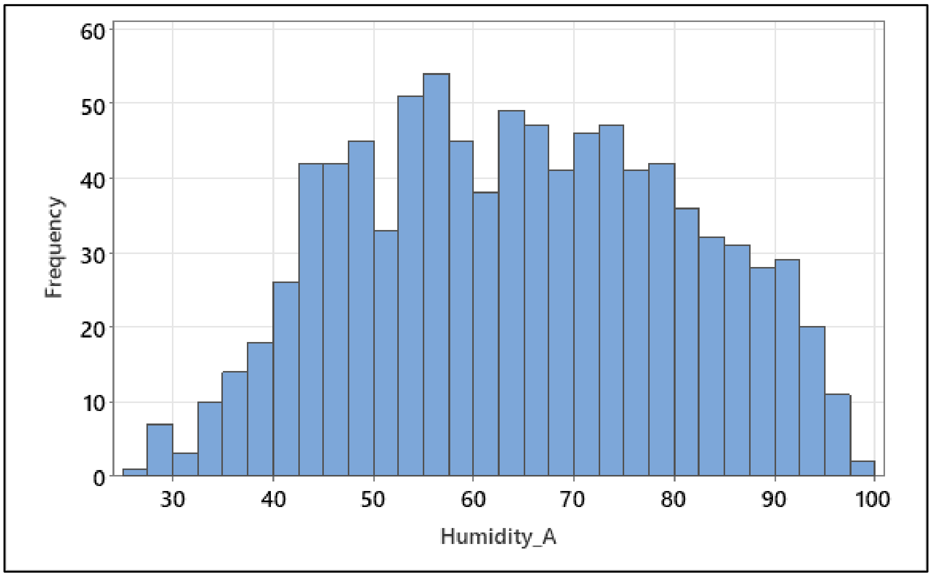


Figure 5: Distribution of Average Daily Winter Humidity

The most effective regression model for precipitation is shown as Equation 1, where $\hat{Y}_P$ is the predicted precipitation (inches), and X_T, X_W, and X_H are the temperature (°F), wind speed (mph), and humidity (%). Although this relationship had a low r-squared of 0.192, its use in the simulation is appropriate because the prediction needs to account for weather parameter dependencies. The residual variation would also be generated randomly within the simulation logic. The response standard deviation of the response prediction was about 0.125, and this variation followed a normal distribution with several outliers (i.e., intense storm days).

$$\hat{Y}_P = 0.24 + 0.015X_T - 0.048X_W - 0.008X_H - 0.00017X_T^2 + 0.0025X_W^2 + 0.000058X_H^2 \quad (1)$$

Part 2 – Clustering

The k-means clustering method was used to categorize hourly patterns into similar groups for each of the five weather parameters. Each of the five standardized data sets consists of a time series vector of length 24, denoted as $y^{(i)} = \{{y_1}^{(i)}, {y_2}^{(i)}, \ldots, {y_t}^{(i)}, \ldots {y_{24}}^{(i)}\}$, where $i = 1, 2, \cdots, 931$ (for each of the 931 days). Each cluster is characterized by its centroid vector, which represents the mean value by hour for each cluster. The number of clusters was determined, in part, by evaluating a curve showing the within cluster sum of squares for each potential number of clusters. The curve's elbow is usually determined visually or mathematically. In many cases, the analyst's goal is to minimize the number of clusters to allow for a more simplified application that uses the various clusters (e.g., market segmentation). In this application, minimizing the number of clusters is not a concern because the simulation can easily handle any number of clusters.

The criterion for choosing the number of clusters was based on their weather modeling accuracy. As illustration, consider Figure 6 that shows six clusters that were generated for the standardized humidity outcomes. Each graph shows the hourly fluctuations across 24 hours for 10 example days in each cluster. Along with the 6 clusters for humidity, the k-means clustering analysis resulted in 6 clusters for temperature, 5 clusters for wind speed, 5 clusters for rain (for rainy days), and 5 clusters for snow (for snowy days).

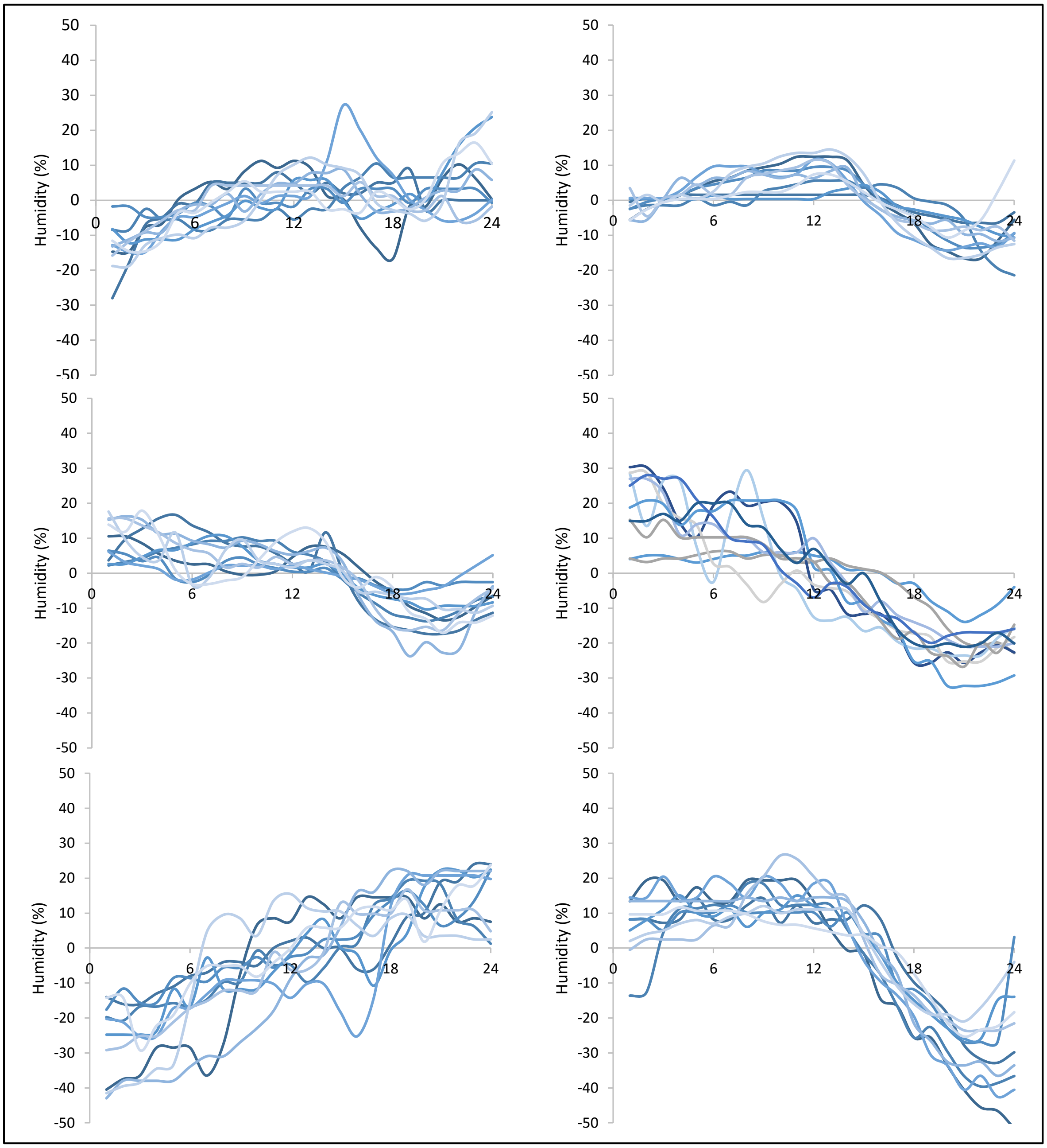


Figure 6: Six Humidity Clusters (by Hour of Day)

Input Model

The simulation's input model generates hourly weather outcomes. Daily weather is generated for average temperature, wind speed, humidity, and total precipitation. The hourly cluster for each of the five weather inputs is chosen randomly based on their likelihoods. This determination must account for any dependencies that exist among the five sets of clusters. An analysis concluded that all clusters were mutually independent, except there was a significant correlation between the snow and humidity clusters, and between the rain and wind speed clusters. Consequently, the generation of the rain and/or snow cluster will be based on a set of probabilities that are conditioned on the identity of the humidity and wind speed clusters.

The input model for generating hourly distributions for each of the five weather parameters includes unconditioned probabilities associated with the hourly temperature, humidity, and wind speed clusters. The likelihoods for each cluster associated with rain were based on the conditional probability for each cluster given the identity of the wind speed cluster, and the likelihoods for each cluster associated with snow were based on the conditional probability for each cluster given the identity of the humidity cluster.

Because temperature will vary over a day, one day can consist of rain only, snow only, or a mix of rain and snow. The input model requires that the type of precipitation for each hour be based on the predicted water content (i.e., precipitation) and the temperature. This determination must be done on an hourly, not daily, basis because one day can include multiple forms of precipitation.

When the hourly temperature is above 34°F, it is assumed that the precipitation takes the form of rain, and the predicted rain amount is equal to the predicted precipitation level. When the hourly temperatures falls below 30°F, it is assumed that the precipitation takes the form of snow, and the predicted snow amount is equal to the product of the predicted precipitation level and the predicted SWR (e.g., if 0.3 inches of precipitation is predicted, the temperature is 28°F, and the SWR is 11.5, the snow prediction is equal to 3.5 inches). When the hourly temperatures falls between 30°F and 34°F, it is assumed that the precipitation takes the form of a snow-rain mix (in equal proportions). In this case, the SWR will likely be very low (i.e., wet snow) and the snow contribution will be lower than it would be for hours with lower temperatures. Figure 7 illustrates this model component for an example 24-hour period.

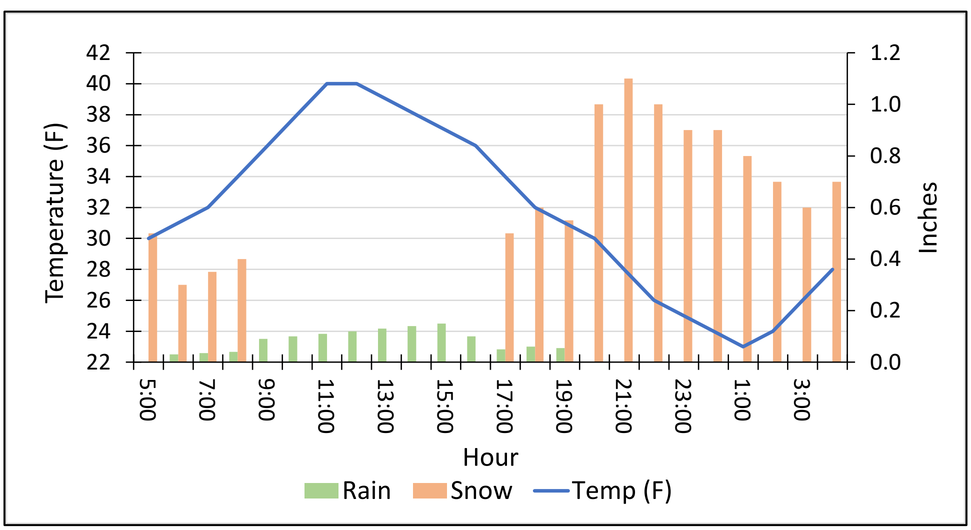


Figure 7: Example Rain-Snow Prediction

EXAMPLE

The simulation can be used to validate severity models for a typical winter day or an extreme day (recall that residuals of the precipitation regression model had outliers representing intense storm days). An example illustrates the implementation of the input model. Consider a winter day that has an average temperature of 20°F, an average wind speed of 15 mph, and an average humidity of 90% (these values would have been generated from the 3-parameter Weibull, gamma, and truncated normal distributions, respectively). For this example, a fairly significant winter storm will be simulated so the total precipitation is input as 0.75 inches. The clusters are generated randomly resulting in the hourly fluctuations shown in Figure 8.

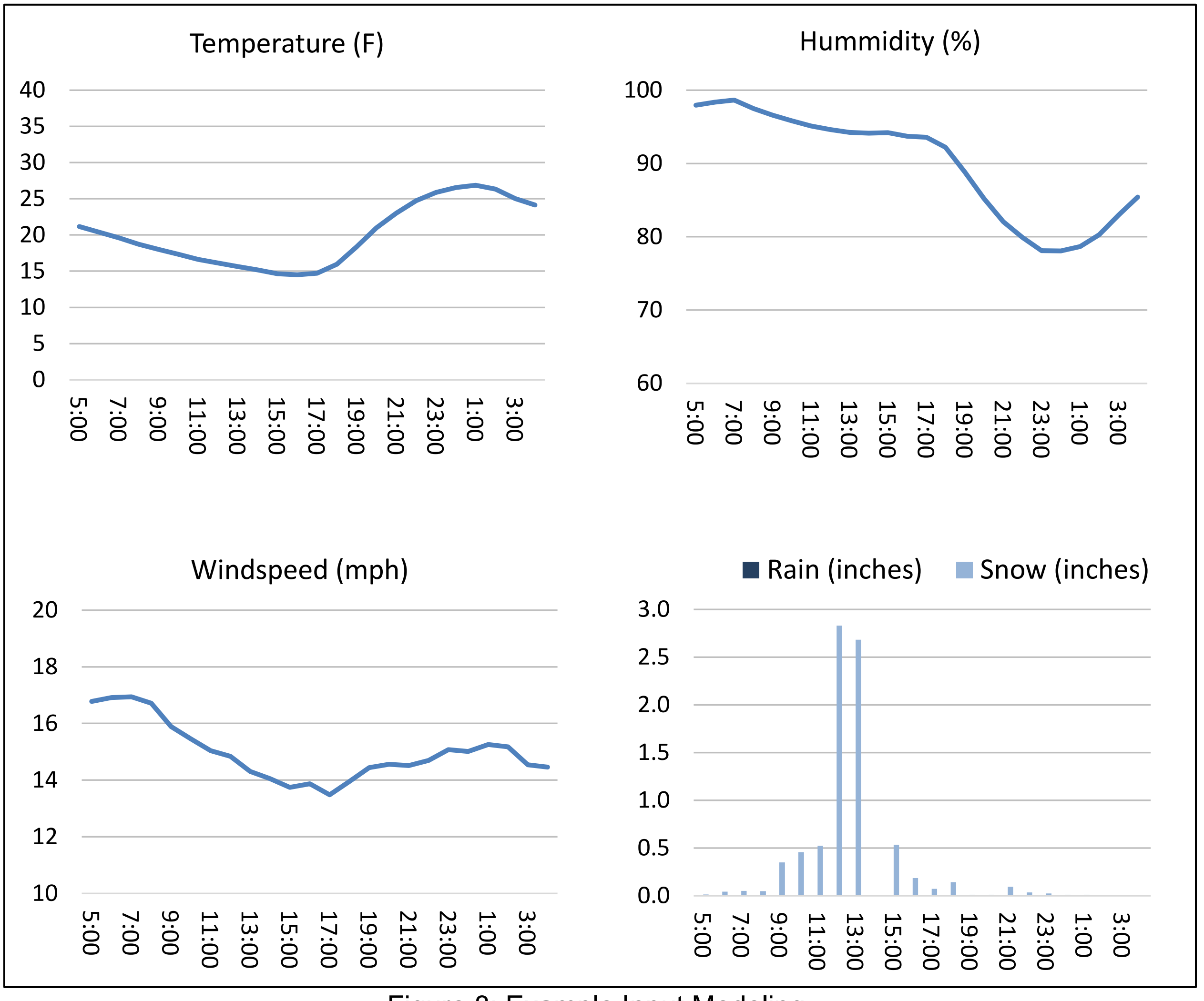


Figure 8: Example Input Modeling

Figure 9 shows the resulting severity predictions for snow depth, slick road probability, snow weight, signal freezing, parking lot area lost, and ingestion risk (which is worse for higher values of SWR). Because their main purpose is to support resource planning, the models assume that no action is taken (e.g., snow depth remains at its peak level). These results, and other results of running the simulation, were reviewed by the MBTA risk and emergency managers. By using the face validation method, they considered the results to be accurate. However, the authors are contemplating more technical approaches to validate the simulation results.

Figure 9: Severity Predictions

STATUS AND FUTURE WORK

Future work can explore several areas to enhance the effectiveness of the current validation effort. Refining and enhancing the machine learning models through further research and analysis will be considered. Incorporating additional datasets or utilizing more advanced machine learning techniques could lead to better input model generators resulting in more effective validation of severity predictions. Once the system validation is complete, integration with other decision-making tools used by MBTA planners and managers would allow for more efficient and comprehensive management of winter storm operations.

Incorporating real-time weather and traffic data into the system will enable more accurate and timely risk assessments for MBTA operations during winter storms. Expanding the scope of the severity models and methodologies to other public transportation systems facing similar challenges in different regions or countries could help improve operational efficiency in a broader context. By adapting the system for use in other locations, the benefits of the predictive models could be shared across multiple transportation networks. Conducting a study to assess the long-term impact of the decision support system on the MBTA's operational efficiency, resource allocation, and overall performance during winter storms would help identify areas for continued improvement.

ACKOWLEDGEMENTS

This work, which commenced in January 2021 as a collaboration with the MBTA and City Innovate, was undertaken by the Boston University Metropolitan College Decision Sciences Laboratory (DSLab). The authors wish to acknowledge the work of other researchers who made significant contributions to the development of the Winter Storm Planning System. The primary contributor was Yiwei (Danielle) Song, Senior Research Associate at the DSLab. Other contributors included Xiaotong Ding, Ruojia Peng, Ching-Wei Kuo, Tracy Ning, and Jennifer Li.

REFERENCES

Back, L.E., & Bretherton, C.S. (2005). The relationship between wind speed and precipitation in the Pacific ITCZ. *Journal of Climate*, *18*(20), 4317-4328.

Barančoková, M., & Kenderessy, P. (2014). Assessment of landslide risk using GIS and statistical methods in Kysuce region. *Ekológia (Bratislava)*, *33*(1), 26-35.

Biller, B., & Ghosh, S. (2006). Multivariate input processes. *Handbooks in Operations Research and Management Science*, *13*, 123-153.

Biller, B. (2009). Copula-based multivariate input models for stochastic simulation. *Operations Research*, *57*(4), 878-892.

Bradley, P.S., & Fayyad, U.M. (1998). Refining initial points for k-means clustering. *Proceedings of the 15th International Conference on Machine Learning*, 91-99.

Bretherton, C.S., Peters, M.E., & Back, L.E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate, 17*(7), 1517-1528.

Carlton, J., Peasgood, T., Mukuria, C., Connell, J., Brazier, J., Ludwig, K., ... & Augustovski, F. (2022). Generation, selection, and face validation of items for a new generic measure of quality of life: the EQ-HWB. *Value in Health*, *25*(4), 512-524.

Gorinevsky, D. (2015). Fault isolation in data-driven multivariate process monitoring. *IEEE Transactions on Control Systems Technology*, *23*(5), 1840-1852.

Guan, L., Yang, J., & Bell, J.M. (2007). Cross-correlations between weather variables in Australia. *Building and Environment*, *42*(3), 1054-1070.

Kilsby, C.G., Jones, P.D., Burton, A., Ford, A.C., Fowler, H.J., Harpham, C., ... & Wilby, R. L. (2007). A daily weather generator for use in climate change studies. *Environmental Modelling & Software*, *22*(12), 1705-1719.

Kumpf, A., Tost, B., Baumgart, M., Riemer, M., Westermann, R., & Rautenhaus, M. (2017). Visualizing confidence in cluster-based ensemble weather forecast analyses. *IEEE Transactions on Visualization and Computer Graphics*, *24*(1), 109-119.

Li, L., Liu, Y., Shi, Z., & Wang, P. (2020). Multivariate sensitivity analysis for dynamic models with both random and random process inputs. *Applied Mathematical Modelling*, *81*, 92-112.

Metsalu, T., & Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Research*, *43*(W1), W566-W570.

Naylor, T.H., & Finger, J.M. (1967). Verification of computer simulation models. *Management Science*, *14*(2), B-92.

Ren, B., & Barnett, I. (2022). Autoregressive mixture models for clustering time series. *Journal of Time Series Analysis*, *43*(6), 918-937.

Richardson C.W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research, 17*(1):182-90.

Sanhudo, L., Rodrigues, J., & Vasconcelos, F.E. (2021). Multivariate time series clustering and forecasting for building energy analysis: Application to weather data quality control. *Journal of Building Engineering*, *35*, 101996.

Sargent, R.G. (2010). Verification and validation of simulation models. *Proceedings of the 2010 Winter Simulation Conference*, 166-183.

Seixas-Mikelus, S.A., Kesavadas, T., Srimathveeravalli, G., Chandrasekhar, R., Wilding, G.E., & Guru, K.A. (2010). Face validation of a novel robotic surgical simulator. *Urology*, *76*(2), 357-360.

Sokal, R.R. (1974). Classification: Purposes, principles, progress, prospects: clustering and other new techniques have changed classificatory principles and practice in many sciences. *Science*, *185*(4157), 1115-1123.

Srikanthan, R., & McMahon, T.A. (2001). Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, *5*(4), 653-670.

Steadman, R.G. (1984). A universal scale of apparent temperature. *Journal of Applied Meteorology and Climatology*, *23*(12), 1674-1687.

Wilks, D.S., & Wilby, R.L. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, *23*(3), 329-357.

Zeng, X. (1999). The relationship among precipitation, cloud-top temperature, and precipitable water over the tropics. *Journal of Climate, 12*(8), 2503-2514.

DECISION SCIENCES INSTITUTE

Weathering the Storm: An Efficiency Analysis Amidst New Jersey Universities' Crisis

Devin Echavarria
New Jersey City University, Jersey City, NJ 07311
Email: dechavarria@njcu.edu

Ruth Ortiz
School of Business, New Jersey City University, Jersey City, NJ 07311
Email: rortiz@njcu.edu

EunSu Lee
Management Department, New Jersey City University, Jersey City, NJ 07311
Email: elee3@njcu.edu

ABSTRACT

Higher education experiences decreasing enrollment and focuses on student success, while concerning efficiency and utilization of limited resources. Data envelopment analysis (DEA) is used to evaluate the relative efficiency of Decision-Making Units (DMUs) with data sourced from the National Center for Education Statistics (NCES) and the Department of Education College Scorecard. The study demonstrated the relative efficiency measure of the four-year academic institutes in New Jersey. The findings from this study offer a strategic roadmap for higher education institutions. Institutions can enhance their operational efficiency and foster student success by making informed decisions based on the results.

KEYWORDS: Relative efficiency, Data envelopment analysis, Linear programming, Benchmarking, and Higher education

INTRODUCTION

New Jersey higher education institutions (HEIs) face unprecedented challenges in the current economic landscape. The ongoing financial strain emphasized by the COVID-19 pandemic has put numerous colleges and universities in precarious situations (Clark, 2023). As these institutions battle a looming enrollment cliff that threatens to plunge them further into crisis, there is a growing imperative to evaluate and enhance their operational and teaching efficiency. The present study seeks to evaluate the operational and teaching efficiency of higher education institutions in New Jersey, with a particular focus on New Jersey City University as a case study.

Utilizing a data envelopment analysis (DEA), we examined private and public 4-year colleges in New Jersey. The aim was to identify potential areas for efficiency improvement within individual institutions, thereby enhancing their overall performance. To measure teaching and operational efficiency, we deployed the Charnes, Cooper, and Rhodes (CCR) model, a widely accepted approach in efficiency analysis (Smithers, 2019).

Our analysis, informed by assumptions about the mature nature of the higher education industry and the availability and stable demand of resources, involves an examination of key inputs (full-time and part-time faculty and in-state tuition) and outputs ("Normal Time" graduation rate and student loan repayment rate two years after graduation). The results of this analysis provided efficiency scores for each DMU, where a score of 1.0 represents maximum efficiency.

In light of our findings, we developed a roadmap for improving NJCU's performance, identifying specific targets for efficiency gains, exploring staffing structures, and suggesting strategies for enhancing student retention rates. This study, therefore, provides valuable insights into the path towards improved performance in higher education, emphasizing practical and implementable strategies for NJCU and similar institutions.

LITERATURE REVIEW

Postsecondary education has been recognized for the opportunities that it has afforded individuals for professional growth and success, contributing to the improvement of economic and social outcomes. Research has shown that higher levels of education have been associated with higher rates of employability as well as higher salaries (Organisation for Economic Co-operation and Development, 2022). As such, the economic returns of these individuals contribute to overall greater economic growth and productivity of society, reinforcing the significance and value of a postsecondary education in today's world.

Understanding the inherent value of a postsecondary degree, Institutions in Higher Education work to identify guidelines and best practices that will contribute to the success of their students. Guided by their mission and values, how an institution defines student success is often determined by many factors related to the achievements and experiences of the student. Defining student success and the various approaches by which many seek to do so has become a topic of much debate in higher education. Nonetheless, the most frequently cited and commonly noted factors of student success pertaining to higher education outcomes are retention and graduation rates (Smithers, 2019; Cuseo, 2006).

With declining enrollment and retention, institutions today are facing many challenges and struggle to identify methods of improvement to ensure the completion of studies and graduation of their students. The National Center for Education and Statistics (NCES) reported that total undergraduate enrollment in degree granting postsecondary institutions in the United States decreased by 9% between 2009 and 2020 (National Center for Education Statistics, 2023). Additionally, the overall retention rates for full-time students saw a decrease from 76.1% in 2019 to 75.7% in 2020 (Weissman, 2021).

Noting difficulties and decline with enrollment of the traditional college student, one literature review sheds light on how higher education is turning their attention to retention. In doing so institutions are also applying a strategical approach on the evaluation of barriers to persistence in higher education for achievement of institutional outcomes (Tillman, 2002). Enrollment contributes to the growth and sustainability of an institution and efforts to identify and address barriers to persistence in higher education are essential to the successful completion of studies.

While retention and graduation have been identified as factors of student success, retention and graduation outcomes have become a tool of measure in evaluating the performance and efficiency of degree granting institutions. Striving to rank among the top performing and most efficient among their peers, institutions looking to improve on higher education outputs are evaluating the input factors of benchmark institutions to ensure student success. One application gaining popularity as a method of evaluation for efficiency in Higher Education Institutions is Data Envelope Analysis (DEA). Identifying the various approaches and many benefits of DEA, one study introduces and discusses how application of this method seeks to not only identify areas of weakness in the performance of the institution but to identify solutions for improvement (Nazarko & Saparauskas, 2014).

While HEIs continue to analyze graduation rates as a factor of student success, many have turned their attention to identifying outcomes associated with the achievement of program completion. While research has evidenced that higher levels of education have been associated with higher levels of employability and salaries, additional observations into factors of student success have revealed that there may in fact, be a link between completion of studies (graduation) and student loan repayment (Kreighbaum, 2018).

With student loans on the rise, completion of programs has become more vital to the success of students not just for job security but for assurance of student loan repayment. In a report by Third Way, an analysis of loan repayment data showed that students who completed their degree were 20 percent more likely to follow through with loan repayments verses those who did not graduate, supporting how completion of program and loan repayment are factors of overall student success (Itzkowitz, 2018).

Collectively, these studies provide valuable insights into the use of DEA in higher education (Kreighbaum, 2018; Itzkowitz, 2018). They help establish a robust theoretical and practical foundation upon which our analysis is built. The focus on teaching efficiency, the improvement of the DEA model, and the consideration of practical and accessible inputs and outputs all contribute to a comprehensive understanding of efficiency in higher education settings.

DATA AND MODEL

Data

The data for this study is sourced from the National Center for Education Statistics (NCES) and the Department of Education College Scorecard, both renowned for their comprehensive and credible datasets pertaining to education in the United States.

The NCES, as the primary federal entity for collecting and analyzing data related to education, accumulates information through an array of strategies including surveys, administrative data collection, and collaborations with state and local education agencies. This ensures a high degree of reliability and comprehensiveness in the data obtained. For this study, specific datasets available in the electronic catalog of NCES products were utilized (National Center for Education Statistics, 2023).

The Department of Education College Scorecard is a valuable resource that provides in-depth information on the cost, graduation rate, employment rate, and average amount borrowed for federal loans among other metrics for colleges and universities across the country. Data for the College Scorecard is collected from a variety of sources, including federal reporting from institutions, data on federal financial aid, and tax information (US Department of Education, 2023).

In transforming this raw data for analysis, several steps were taken. First, the data was cleaned to eliminate inconsistencies and errors, ensuring that all information was accurate and applicable. Secondly, the data was normalized. Normalization adjusted the values measured on different scales to a notionally common scale, allowing for direct comparison of various data points. Lastly, the data was disaggregated where necessary to provide a sufficient level of detail. Disaggregation involved breaking down the data into smaller categories or segments, providing a more granular view that could better inform the DEA analysis.

The efficiency analysis was conducted using R Studios and packages DEAr, rDEA, Benchmarking, and Pysch. This process ensured that the data used was both reliable and tailored to the requirements of this study, paving the way for a thorough and insightful efficiency analysis.

Linear Programming

Data Envelopment Analysis (DEA), introduced by Charnes, Cooper, and Rhodes (CCR), is a mathematical programming technique used to evaluate the relative efficiency of Decision-Making Units (DMUs). Each DMU consumes certain amounts of inputs to produce certain amounts of outputs, and efficiency is evaluated by comparing the ratio of weighted outputs to weighted inputs across DMUs (Panwar, Olfati, Pant, & Snasel, 2022). The mathematical representation of a basic CCR DEA model for a single DMU (DMU_i) is as follows:

The efficiency of *i*-th DMU E_i is defined as the ratio of the weighted sum of outputs to the weighted sum of inputs, which can be written as (1):

$$E_i = \frac{\sum_{j=1}^{n_O} O_{ij} w_j}{\sum_{j=1}^{n_I} I_{ij} v_j} \tag{1}$$

Here, O_{ij} is the amount of output j produced by DMU_i, w_j is the weight for output j, and n_O is the number of output variables (i.e., n_O=2). I_{ij} is the amount of input j consumed by DMU_i, v_j is the weight for input j and n_I is the number of output variables (i.e., n_I=3).

The weights are determined to maximize the efficiency of DMU_i, which leads to the following objective function (2):

$$MAX\colon \sum_{j=1}^{n_O} O_{ij} w_j \tag{2}$$

This objective function aims to maximize the weighted sum of outputs for DMU_i. There are, however, three constraints that must be considered in this linear programming problem. The first constraint ensures that no DMU can have an efficiency score greater than 1, which is written as (3):

$$\sum_{j=1}^{n_O} O_{kj} w_j \le \sum_{j=1}^{n_I} I_{kj} v_j \,, \forall\; DMU\; k = \{1,2 \ldots, K\} \tag{3}$$

This means that the weighted sum of outputs for any DMU (DMU_k) cannot exceed its weighted sum of inputs. The second constraint ensures that the weighted sum of inputs for DMU_i is equal to 1, which can be written as (4):

$$\sum_{j=1}^{n_1} I_{ij} v_j = 1 \tag{4}$$

This constraint is crucial for making the efficiency scores of different DMUs comparable to each other. The final constraint ensures that all weights are non-negative (5):

$$w_j, v_j \geq 0, \forall\, j = \{1,2,\ldots,J\} \tag{5}$$

This is a standard constraint in DEA models, ensuring that neither inputs nor outputs are given negative importance in the efficiency calculations. These equations collectively represent a fractional linear programming problem, and they form the basis of the CCR DEA model for measuring the relative efficiency of DMUs (Ragsdale, 2017).

Choice of Factors

The selection of input and output variables is central to the accuracy and relevancy of a DEA study. For this study, our goal is to measure the teaching and operational efficiency of New Jersey Universities in relation to student success. This goal informed our choice of input and output variables.

Inputs:

1. Number of Full-Time (FT) Faculty (X1): This variable was chosen as a measure of the academic resources available at each institution. Having a higher number of full-time faculty is generally associated with more individual attention for students, and more faculty time devoted to teaching and mentoring.
2. Number of Part-Time (PT) Faculty (X2): This measure provides a sense of the flexible academic capacity of an institution. Part-time faculty can supplement the teaching resources provided by full-time faculty, but they may also indicate reliance on non-permanent staff, which could impact the continuity of student learning (Kuah & Wong, 2011).
3. In-State Tuition in $ (X3): This measure serves as a proxy for the financial resources available to the institution and the financial burden on the students. Lower tuition can often lead to increased access for in-state students but might also indicate lower institutional resources if not supplemented by out-of-state or international tuition fees or other forms of funding.

Outputs:

1. "Normal Time" Graduation Rate (Y1): As defined by the NCES, the "Normal time" graduation rate is 150% of the time necessary to complete the respective degree. For example, the "normal time" for a 4-year degree will be measured by the 6-year graduation rate. This is a commonly used measure of student success in higher education studies. A higher graduation rate indicates that the university is effectively supporting its students towards degree completion within a reasonable time frame (National Center for Education Statistics, 2023).
2. Student Loan Repayment Rate 2 Years After Graduation (Y2): This measure offers insight into the financial success of graduates, which can be influenced by the quality of

education and support they received during their time at the university. A higher repayment rate indicates that graduates are finding employment and earning enough to start repaying their student loans, which is a positive outcome for students and universities alike.

This mix of inputs and outputs allows us to examine the resource utilization of universities (inputs) and the outcomes achieved (outputs). In the context of our study, these variables are directly aligned with our goal of measuring the teaching/operational efficiency of New Jersey Universities for student success (Table 1). In the next stage of our study, we apply these factors in DEA.

	Table 1: Data for 28 New Jersey Universities					
DMU	Universities	FT Faculty	PT Faculty	In-State Tuition	Graduation Rate	2Yr Loan Repayment Rate
		X1	X2	X3	Y1	Y2
1	New Jersey Institute of Technology	468	354	26874	74%	38%
2	Rutgers, Camden	305	335	22868	71%	31%
3	Rutgers, New Brunswick	3435	2260	22994	84%	31%
4	Rutgers, Newark	584	388	22493	65%	31%
5	Kean University	353	991	17025	48%	22%
6	Montclair State University	614	1250	16433	67%	29%
7	New Jersey City University	237	464	19850	39%	22%
8	Ramapo College of New Jersey	207	261	22406	72%	35%
9	Rowan University	831	1015	21589	69%	33%
10	The College of New Jersey	372	444	23421	87%	44%
11	The Richard Stockton College of NJ	329	406	20439	76%	30%
12	William Paterson University of NJ	353	475	20160	56%	25%
13	Bloomfield College	49	129	34197	38%	11%
14	Caldwell University	76	198	43500	59%	24%
15	Centenary University	64	111	40012	68%	31%
16	Drew University	131	128	46702	74%	28%
17	Fairleigh Dickinson University, Florham	164	239	49116	62%	27%
18	Fairleigh Dickinson University, Metropolitan	139	325	46616	54%	25%
19	Felician University	87	168	40960	52%	23%
20	Georgian Court University	92	170	39540	58%	25%
21	Monmouth University	297	384	46007	69%	33%
22	Pillar College	12	91	28656	35%	10%
23	Princeton University	1006	193	49552	98%	39%
24	Rider University	185	280	49860	61%	33%
25	Saint Elizabeth University	57	159	41084	39%	22%
26	Saint Peter's University	104	181	43110	58%	19%
27	Seton Hall University	461	517	49990	72%	31%
28	Stevens Institute of Technology	308	153	58702	87%	45%

The variables X1 and X2, representing the input measures, display a high level of dispersion, as evidenced by their relatively large standard deviations of 640.8 and 458, respectively (Table 2). In both cases, the mean exceeds the median, indicating a positive skew or a rightward distribution tail. This skewness suggests the presence of a few universities with unusually high values for these measures, raising the average above the middle value. Here too, the mean exceeds the median, implying the presence of outliers with high values.

Y1 and Y2, likely representing output measures, exhibit smaller dispersion levels with standard deviations of 0.157 and 0.082, respectively. For these variables, the mean is slightly lower than the median, suggesting a negative skew or leftward distribution tail. This indicates a few universities with unusually low output values. These skewness characteristics and variability levels should be considered when conducting any subsequent analyses or when interpreting model results.

Table 2: Descriptive Statistics

Variables	Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum	Std. Dev.
X1	12	101	267	404.3	394.2	3435	640.8
X2	91	169.5	302.5	431	449	2260	458
X3	16433	22471	36869	34434	46159	58702	12788.2
Y1	0.350	0.555	0.660	0.640	0.725	0.980	0.157
Y2	0.100	0.238	0.295	0.285	0.330	0.450	0.082

Efficiency Analysis

This analysis employs a DEA model developed by Charnes, Cooper, and Rhodes (CCR), which has proven to be an effective and flexible tool in assessing efficiency, especially in contexts like ours where multiple inputs and outputs are at play.

The CCR model was selected because it assumes a constant return to scale, which aligns with our assumptions about the operation of the institutions being analyzed. We assumed that our Decision-Making Units (DMUs) – in this case, the 28 New Jersey universities – are operating at an optimal scale. This reasoning is based on several factors. Firstly, these universities represent mature industries where optimal scales of operation have been identified and implemented over years of operation. Secondly, like for-profit businesses, these non-profits need resources to operate effectively, and the efficient use of these resources can lead to better outcomes in achieving their mission. Lastly, the demand for higher education is relatively stable and predictable, allowing universities to adjust their scale of operation to meet this demand efficiently.

This framework was applied to publicly available data for the universities. For example, one of the universities (DMU 1) had 468 full-time faculty members (X1), 354 part-time faculty members (X2), and an in-state tuition of $26,874 (X3). Its “Normal time” graduation rate was 74% (Y1) and the student loan repayment rate 2 years after graduation was 38% (Y2). Using the CCR model, the efficiency score for this university was determined to be 0.869.

Our efficiency analysis provides a clear path for evaluating and improving the efficiency of universities at the institutional level. Our methodology and assumptions, combined with our chosen input-output mix, enable us to provide meaningful and actionable insights into operational efficiency in New Jersey's higher education sector.

RESULTS

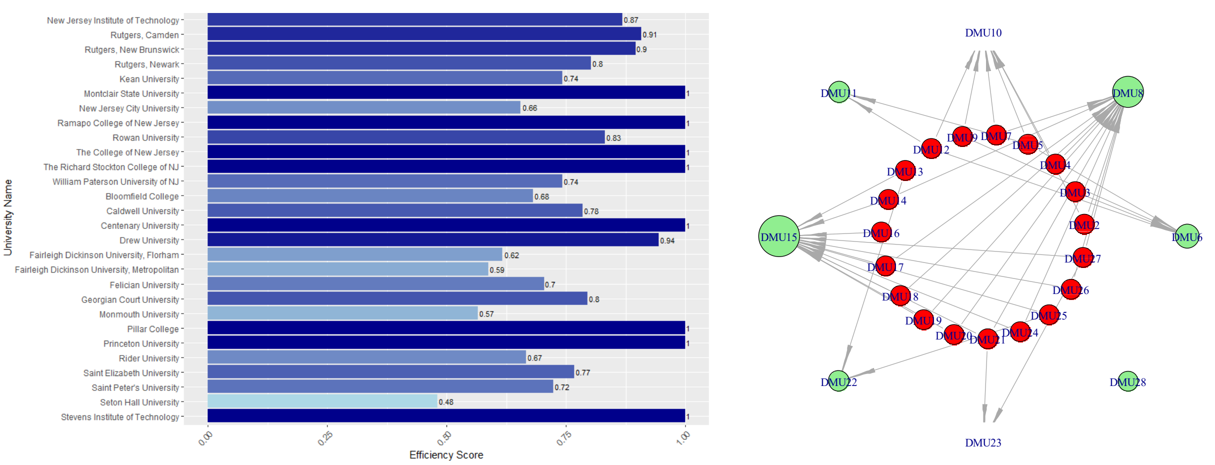
The analysis resulted in a range of efficiency scores for the 28 Decision Making Units (DMUs), reflecting the performance of New Jersey universities in achieving student success with their available resources.

Analysis

In Figure 1a, the CCR efficiency scores ranged from 0.481 to 1.000, with a mean efficiency score of approximately 0.842. The variation in scores indicates a broad spread of efficiency across the 28 universities. For example, a closer examination of DMU 7, New Jersey City University (NJCU), reveals an efficiency score of 0.656, which is below the average efficiency score for all universities. When compared to similar universities, defined here as non-flagship public universities (DMUs 5-12), NJCU's efficiency score places it in the lower range, reflecting a need for improvement.

Figure 1: Efficiency Scores for New Jersey Universities

(a) Efficiency scores (b) Benchmarking units



Note: Figure 1(b) was generated by using 'deaR' package

Highlighting DMU 7 (New Jersey City University)

NJCU, with its CCR score of 0.656, ranks 25th out of the 28 universities evaluated, suggesting considerable room for improvement. Within its peer group of non-flagship public universities (DMUs 5-12), NJCU ranks 8th out of 8, reinforcing this observation.

The universities that serve as a benchmark for NJCU (with a CCR score of 1.000) are DMUs 8 and 10, or Ramapo College of New Jersey and The College of New Jersey, respectively. These institutions could serve as a model for NJCU to identify best practices and strategies to enhance their efficiency.

To improve the efficiency scores of the universities, it's crucial to understand the sources of inefficiency. This involves comparing the input/output mix data, identifying targets for improvement, and addressing areas of slack.

1. Comparing Input/Output Mix Data: This involves looking at the differences in the data between inefficient units and their efficient counterparts (benchmark DMUs). By doing so, we can identify which inputs and outputs need attention.
2. Identifying Targets: Once the sources of inefficiency are recognized, specific targets can be set for improving performance. For instance, if an inefficient DMU has an excess of part-time faculty (input), the university may set a target to reduce this number and increase full-time faculty. Conversely, if the 6-year graduation rate (output) is low, a target could be set to improve this metric.
3. Addressing Slacks: Slack refers to the amount of input that could be reduced without affecting the output, or the amount of output that could be increased without needing more input. In terms of efficiency score, a university with a score of 0.80 has a slack of 0.20. This indicates an inefficiency of 20%, meaning there's potential to reduce inputs or increase outputs to improve efficiency.

For inputs, the goal should be to reduce slack, effectively using resources more efficiently. For outputs, the goal should be to increase them, thereby enhancing the university's effectiveness in achieving student success.

By following this approach, universities can systematically identify and address their sources of inefficiency, moving them closer to their benchmark DMUs and improving their overall efficiency scores.

In conclusion, our efficiency analysis offers strategic insights for improving the performance of New Jersey's higher education institutions. With a specific focus on New Jersey City University, our study provides a roadmap for enhancing operational and teaching efficiency to promote student success.

DISCUSSION

Our efficiency analysis has offered valuable insights into the operational and teaching efficiency of New Jersey's higher education institutions. Using these insights, we focus on New Jersey City University (NJCU or DMU 7), offering key recommendations to enhance its efficiency and foster student success.

1. Improvement of Graduation and Loan Repayment Rates:

According to the slacks from the CCR scores, NJCU could significantly improve its efficiency by targeting a 6-year graduation rate of 52% (a rise of 13%) and a student loan repayment rate of 29% (an increase of 7%). Achieving these targets, while maintaining current input levels, would align NJCU's efficiency with that of its benchmark Decision-Making Units (DMUs).

2. Increase in Full-time Faculty:

Our analysis shows that NJCU's percentage of full-time faculty is significantly lower than other universities of similar size. Benchmark DMUs (8 and 10), namely Ramapo College of New Jersey and The College of New Jersey, have full-time faculty percentages exceeding 43%, whereas NJCU's rate stands at 33%. Prioritizing the hiring of more full-time staff could offer several advantages, such as:

- Consistency and Stability: More full-time staff would ensure greater consistency in course offerings and stable program administration.
- Faculty Engagement and Governance: Full-time faculty members are typically more involved in faculty governance and campus life, fostering a stronger and more vibrant academic community.

The roadmap from our analysis provides a clear direction for enhancing NJCU's efficiency and promoting student success. By following these recommendations, NJCU can move closer to its benchmark DMUs, thereby significantly improving its performance.

CONCLUSION

This study's utilization of the Data Envelopment Analysis (DEA) model has added valuable depth and dimension to the understanding of efficiency in higher education, especially in New Jersey's context. Though well-established, the DEA models continue to provide novel insights into diverse fields and offer opportunities for further explorations and enhancements.

Our analysis, which employs a one-stage DEA model, marks a significant contribution to this expansive body of work. It bridges a gap in literature by offering an approach that provides granular insights, helping to guide targeted efficiency improvements in higher education institutions. We've applied this model to both public and private universities in New Jersey, focusing our analysis on a range of input and output variables.

The findings from this study offer a strategic roadmap for higher education institutions, such as New Jersey City University. These institutions can enhance their operational efficiency and foster student success by making informed decisions based on our results.

In essence, our study demonstrates the versatility and efficacy of the DEA model in assessing and improving efficiency within higher education. It serves as a beacon for future research and acts as a catalyst for institutions striving for higher operational efficiency.

Limitation of this Study

While this study provides valuable insights into the operational efficiency of higher education institutions in New Jersey, it is not without limitations, most of which arise from the constraints associated with the available data and inherent limitations of the DEA model.

1. Limited Access to Data: Our research relied heavily on publicly available data, which may not encompass all the aspects necessary for a comprehensive efficiency analysis. Some potentially important variables might not have been included due to lack of access to private or undisclosed data. Examples could include specific budgetary details, student demographic characteristics, faculty qualifications, and more detailed measures of student success. The absence of these data points may influence the precision of our efficiency scores.
2. Use of a Single Year of Data: Our analysis is based on a single year of data, which might not capture the dynamic nature of efficiency in the context of higher education. Efficiency may change over time due to changes in strategy, policies, and external influences. Therefore, our efficiency scores should be interpreted as a snapshot in time, not necessarily indicative of long-term performance.

3. Subjectivity in Input and Output Selection: Despite our best efforts to choose relevant and appropriate input-output combinations, this selection process is inherently subjective. Other researchers might choose different variables to represent inputs and outputs, which could yield different efficiency scores.

These limitations should be taken into account when interpreting the findings of this study and should guide the directions for future research in this area.

Future Research

Retention of part-time students at NJCU is a critical area requiring attention. The university has a retention rate of 33% from year one to year two for part-time students. Comparatively, the benchmark DMUs both boast a part-time student retention rate of 100%. To improve part-time student retention, NJCU could focus on:

- Flexible Scheduling and Course Delivery: Courses offered in the evenings, on weekends, or online could better accommodate part-time students' other commitments.
- Targeted Support Services: Offering services such as academic advising, career services, and other support tailored to part-time students' schedules can help them overcome challenges to their success.
- Financial Aid and Scholarships: Expanding financial aid opportunities for part-time students could alleviate their financial stress.
- Building Community: Creating opportunities for part-time students to engage with peers and the wider campus community could enhance their sense of belonging and commitment to their studies.
- Recognition of Prior Learning: Acknowledging the skills and knowledge that students have gained through work or other experiences can accelerate their progress toward their degree.

In conclusion, the study provides a starting point for a more in-depth exploration into the operational efficiency of higher education institutions. The proposed future research directions could enhance our understanding of the factors affecting university efficiency, provide robustness to our findings, and offer additional insights to inform policy and practice in the higher education sector.

REFERENCES

Abing, S., Barton, M., Dumdum, M., Bongo, M., & Ocampo, L. (2018). Shapley value-based multi-objective data envelopment analysis application for assessing academic efficiency of university departments. *Journal of Industrial Engineering International*, 733-746.

Agapie, A., Badin, L., Dedu, S., Giuclea, M., Miricioi (Valimaranu), H., & Solomon, O. (2020). Analyzing Factors Accountable for Success in Completing University Studies. *Review of International Comparative Management*, 423-432.

Alabdulmenem, F. (2016). Measuring the efficiency of public universities: Using Data Envelopment Analysis (DEA) to examine public universities in. *International Education Studies*, 137.

Beaseley, J. (1990). Comparing University departments. *Omega*, 171-183.

Bougnol, M., & Dula, J. (2006). Validating DEA as a ranking tool; An application of DEA to assess performance in higher education. *Ann Oper Res*, 339-365.

Clark, A. (2023, March 17). *Are more N.J. colleges doomed to collapse? 'These are very, very traumatic time,' official says.* Retrieved from NJ.com True Jersey: https://www.nj.com/education/2023/03/are-more-nj-colleges-doomed-to-collapse-these-are-very-very-traumatic-times-official-says.html

Cuseo, J. (2006). Student Success: Definition, Outcomes, Principles and Practicies. *Esource for College Transitions*.

Figueiredo de Franca, J., & Figueiredo, J. (2010). A DEA Methodolgy to evaluate the impact of information asymmetry on the efficiency of not-for-profit organizations with an application to higher education in Brazil. *Ann Oper Res*, 39-56.

Itzkowitz, M. (2018, August 18). *Want More Students To Pay Down Their Loans? Help Them Graduate.* Retrieved from Third Way: https://www.thirdway.org/report/want-more-students-to-pay-down-their-loans-help-them-graduate

Kim, M., Lee, I., & Oh, S. (2016). Measuring Efficiences of Higher Education Using DEA. *International Journal of u- and e- Service, Science, and Technology*, 321-328.

Kreighbaum, A. (2018, August 07). *The Link Between Completion and Loan Repayment*. Retrieved from Inside Higher Ed: https://www.insidehighered.com/news/2018/08/08/link-between-college-completion-and-student-loan-repayment

Kuah, C., & Wong, K. (2011). Efficiency assessment of universities through data envelopment analysis. *Procedia Computer Science*, 499-506.

National Center for Education Statistics. (2023). *Electronic catalog of NCES products (National Center for Education Statistics). publications and Data Products.* Retrieved from National Center for Education Statistics (NCES) - Data & Tools - Most Popular tools: https://nces.ed.gov/collegenavigator/?s=NJ&l=93&ct=2&ic=1&pg=2&cx=1

Nazarko, J., & Saparauskas, J. (2014). Application of DEA Method in Efficiency Evaluation of Public Higher Education Institutions. *Technological and Economic Development of Economy*, 25-44.

Organisation for Economic Co-operation and Development. (2022). *Economic & Social outcomes*. Retrieved from Education GPS.

Panwar, A., Olfati, M., Pant, M., & Snasel, V. (2022). A Review on the 40 Years of Existence of Data Envelopment Analysis Models: Historic Development and Current Trends. *Archives of Computation Methods in Engineering*, 5397-5426.

Ragsdale, C. (2017). *Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Business Analytics 8th Edition.* Cengage Learning.

Salas-Velasco, M. (2020). The technical efficiency of performance of the higher education systems based on data envelopment analysis with an illustration for the Spanish case. *Educational Research for Policy and Practice*, 159-180.

Smithers, L. E. (2019). How Should Institutions of Higher Education Define and Measure Student Success? Student Success as Liberal Education Escapes Definition and Measurement. *Contested Issues in Troubled Times: Student Affairs Dialogues About Equity, Civility, and Safety*, 127-137.

Tillman, C. (2002). *Barriers to Student Persistence in Higher Education.* Wesleyan-Holiness Digital Library.

US Department of Education. (2023). *Data Home: College Scorecard*. Retrieved from Data Home | College Scorecard: https://collegescorecard.ed.gov/data

Weissman, S. (2021, September 14). *Not Coming, Not Staying*. Retrieved from Inside Higher Ed: https://www.insidehighered.com/news/2021/09/15/new-federal-data-confirm-enrollment-declines

DECISION SCIENCES INSTITUTE

What Makes A Company The Best Place to Work?

Tuncay Bayrak
Western New England University
College of Business
Springfield, MA.
tbayrak@wne.edu

ABSTRACT

The aim of this study is to identify the most important attributes and characteristics of Fortune magazine's "100 Best Places to Work for" in 2021. To do so, we performed text analytics on the comments made by their employees describing their working environments. While there is no single set of qualities and attributes that makes a workplace a great workplace to work for, through text mining some common attributes can automatically be identified and extracted from the comments made by the employees.

KEYWORDS: Workplace, employee, company, work

INTRODUCTION

Working environments and the way employees are treated certainly affect employees' morale and performance. Therefore, employees are increasingly interested in working at companies where they are treated with respect and dignity, and their opinions are valued and appreciated. In addition, they are interested in a workplace where they have the necessary tools and management support to do their jobs well.

In the COVID-19 pandemic era, it's more important than ever to create a workplace culture where the executive team pays close attention to the health and well-being of the employees, is open about the challenges facing the business, maintains constant communication with their employees, and both employees and leadership trust one another. Therefore, to address employees' core needs and to be recognized as a great place to work, companies across various industries are developing and creating working environments where open communication and feedback are encouraged.

While there is no single set of characteristics shared by the best places to work, one might wonder what attributes make a company a great place to work. Feffer (2015) argues that the foundation of a great workplace lies in a culture of trust and engagement that unites management and the workforce in a common vision. Therefore, as pointed out by Schwartz (2011), companies must shift the focus from trying to get more out of people, to investing more in them by addressing their core needs such as physical, emotional, mental, and spiritual.
This study examines Fortune magazine's "100 Best Places to Work for" in America in 2021. To identify characteristics and attributes describing the best places to work, we performed text analytics on the comments and statements made by their employees about their working environments, social and physical conditions, scheduling flexibility, and how they were treated during the pandemic.

LITERATURE REVIEW

A number of studies have been conducted in the past to examine the qualities and attributes that make a workplace the best workplace to work. For instance, Filbeck (2001) examined 12 publicly held firms that were ranked as “Better Places to Work” in terms of charitable giving, fair labor practices, progressive benefits, sound environmental practices, and employee satisfaction. A similar study was conducted by Romero (2004) who explored the actual performance of the 1998 “100 Best Companies to Work for in America,” as listed by Fortune Magazine. Having analyzed both stock market data and accounting data to assess firm performance, the study revealed that positive employee relations were beneficial for companies.

Filbeck and Preece (2003) carried out a study to examine the market reaction to the announcement by Fortune of the ‘Best 100 Companies to Work for in America.’ The study asked employees to rate their firms based on several criteria including trust in management, pride in work/company, and camaraderie. The authors found a statistically significant positive response to the announcement of the ‘100 Best Companies to Work for’ by Fortune. Using data from both the American Customer Satisfaction Index (ACSI) and Fortune Magazine's lists of Best Companies, Simon and Devaro (2006) examined the relationship between making the ‘100 Best’ list and customer satisfaction. Based on a subset of the 100 Best in each year from 1994 to 2002, they found strong evidence that firms on the list earn higher customer satisfaction ratings than firms not on the list.

Hinkin and Tracey (2010) analyzed the best companies to work in terms of a culture that emphasizes the value of people, management policies, creative staffing practices, scheduling flexibility, people-oriented training programs, and compensation policies that reflect the organization’s values and link pay to performance. A study by (Butler et al., 2016) explored the relationship between being a “great place to work” and firm performance. The study focused on being a trustworthy employer, a place where workers take pride in their work and enjoy the people with whom they work. In the same year, DeMaria (2016) conducted a study to evaluate internal communications in companies deemed "Best Places to Work." Interviews held by the author with professionals from the "Best Places to Work" listing revealed a close connection between internal communications and corporate culture. The study found that providing an environment where open communication is encouraged reflects the corporate culture and aids in its development. Cycyota et al., (2016) reviewed publicly available documents of Fortune's “100 Best Companies to Work For” ranking to gain insight into how these best companies practice employee volunteerism and whether they link employee volunteerism to their corporate social responsibility strategy. Their findings suggest that many highly regarded companies specifically link employee volunteerism to their corporate social responsibility strategy.

Recently, Vittal (2021) analyzed 11 articles covered between 2003 and 2014 that directly studied the work on Best Companies survey established by the “Great Place To Work Institute.” The author observed that literature on Great Place To Work is segmented into two themes: (a) themes that cover positive consequences of being among the Best Companies such as Customer Satisfaction, Financial Performance and (b) themes that highlight the lessons from the best practices of “100 Best Companies.” Having analyzed Fortune magazine’s 100 Great Places to Work For listing, Henderson (2021) concluded that working environments are more satisfactory if the organization encouraged closeness, camaraderie, and trust among its employees. Organizations that encourage group processes appeared to be considered superior places to work. This analysis showed that group characteristics within organizations have a strong influence on how the organization is perceived by its employees.

While the majority of the aforementioned studies examined Fortune magazine's 100 Best Places to Work from the viewpoint of the employees, a couple of studies took a slightly different approach to exploring such companies. For instance, Bernardi et al., (2006) investigated whether a higher representation of women on a board is an important characteristic necessary to establish the firm on Fortune's "100 Best Companies to Work For" list. They found a positive correlation between the number of female directors and a company's appearance on the "100 Best Companies to Work For" list. Carvalho and Area (2016) ran a study to explore the resilience of the "100 Best Companies to Work for in America" in times of financial crisis by analyzing their long-term financial performance. The authors concluded that the best places to work are resilient in times of crisis as their financial performance is not affected in times of financial crisis.

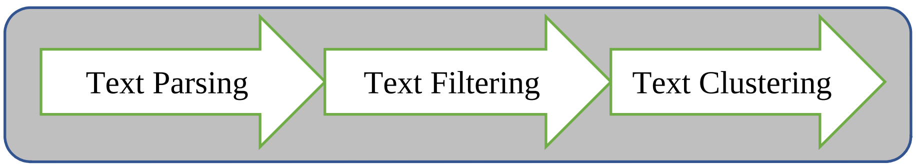
METHODS

Partnering with the Great Place to Work Institute, Fortune magazine published the "100 Best Workplaces to Work for" list in 2021 (Fortune-a, 2021). As summarized by Fortune magazine on its website (Fortune-b, 2021), the Fortune 100 Best Companies to Work For award is based on an analysis of survey responses from more than half a million current employees across the U.S. Great Place to Work selects winners of the Best Workplaces lists primarily based on employees' responses to their survey on issues including how trustworthy, caring, and fair the company is in times of crises; employees' physical, emotional, and financial health; and the company's broader community impact.

Particular attention was paid to how employees' experiences varied depending on their job role, gender, race/ethnicity, payroll status, and other characteristics to ensure that the company is creating a great workplace for all. Companies need to employ at least 1,000 U.S. employees to be considered for the 100 Best Companies list.

The list contains information about each company in terms of their ranks, industries, locations, and what employees are saying about their workplaces. We paid particular attention to the section "What employees are saying," which contains comments and statements made by their employees describing their working environments. After compiling the comments and statements made by the employees of the best workplaces to work for and saving them in a .txt file, we performed text mining on it using SAS Enterprise Miner, an advanced data mining tool from the SAS Institute. The text mining process employed in this study is illustrated in Figure 1.

Figure 1: Text mining process

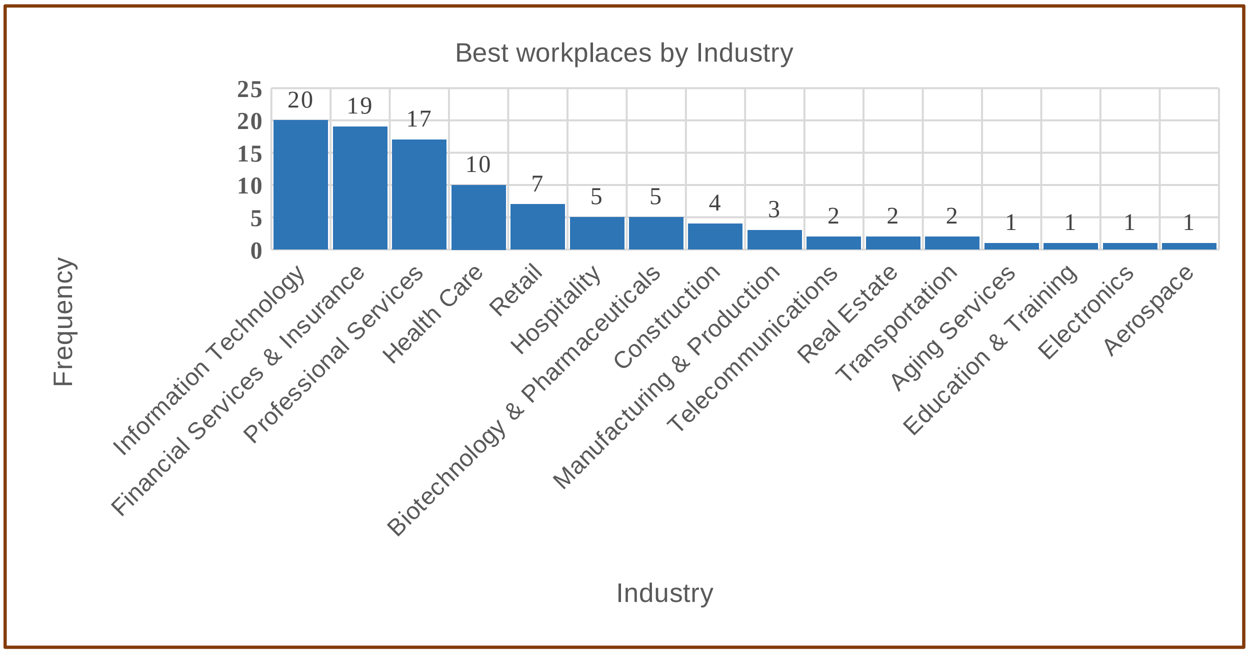


RESULTS

100 Best Places to Work for

Figure 2 summarizes the industries in which the companies included in Fortune magazine's "100 Best Places to Work for" list operate. As seen, the IT industry has the highest number of companies recognized as the best workplaces to work for. Financial services, professional services, and healthcare industries are home to a significant number of companies included in the same list. Such industries as Education, Electronic, and Aerospace are home to the fewest number of companies featured in the same list.

Figure 2: Best workplaces to work for by industry.

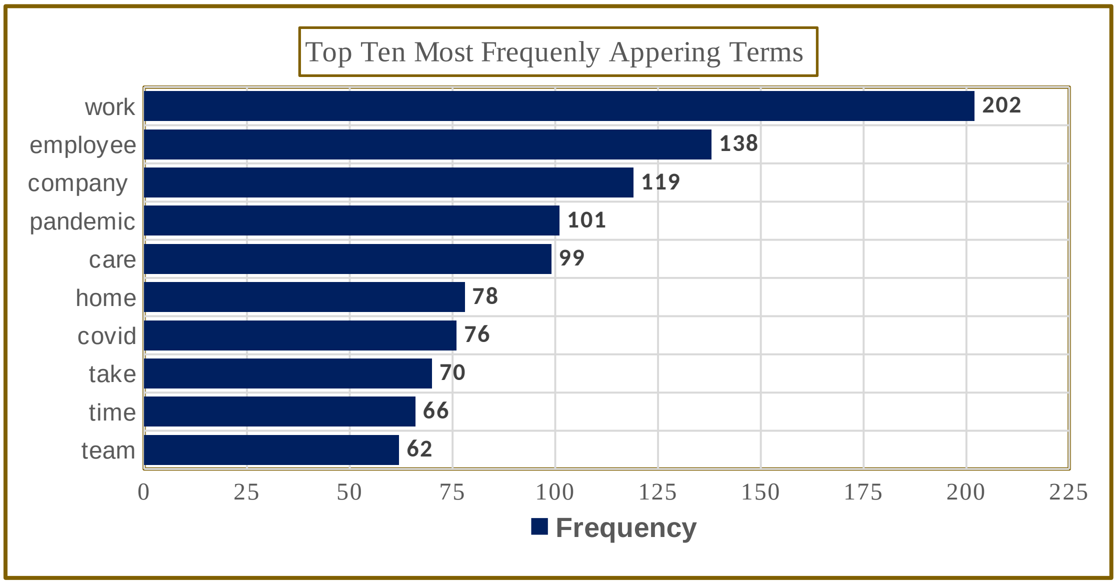


Parsing the Employees' Comments

Much information can be uncovered about the nature and themes of the statements made by the employees describing their workplaces. Therefore, using the Text Parsing node, we first deconstructed and broke down the document containing the comments made by the employees. The Text Parsing node is often used to quantify key information about the terms within a document for further analysis. It identifies and quantifies the most frequently occurring terms in a document collection.

As illustrated in Figure 3, some of the terms that had the highest frequencies are "work", "employee", "company", "pandemic", and "care" appearing 202, 138, 119, 101, and 99 times respectively. In addition, not surprisingly, employees used the terms "pandemic," and "covid" quite frequently when describing the challenges that they faced in their workplaces during the pandemic.

Figure 3: Top ten most frequently occurring terms.



.

By looking at these terms/words one can develop an understanding of what is being talked about in the document containing the statements made by the employees.

Identifying the Most Important Terms

While the text parsing node may aid in transforming unstructured textual data into structured data for easy analysis and quantifying the most commonly occurring key terms and words, it does not necessarily identify the most significant terms. Therefore, having identified the most frequently occurring terms across the employees' statements, we then ran the Text Filter node, which allows one to look through documents and decide which terms have enough significance in the text analysis to keep in subsequent nodes and which terms have less importance and should be dropped from subsequent nodes. In other words, the text filter node is often employed to identify the most significant and relevant terms in a document collection.

The weight of any given term in a document is determined by:

TF-IDF = Term Frequency (TF) * Inverse Document Frequency (IDF).
tf-idf(t, d) = tf(t, d) * log(N/(df + 1))
tf(t,d) = the number of times a term t occurs in document d
df(t) = The number of documents in the collection in which a term t occurs.
where N is the number of documents in the collection (Nguyen, 2014).

Table 1 summarizes the top ten most important terms along with their weights. As seen, the terms with significance include such important terms as "financial", "customer", "team" and "pay" with weights of 0.583, 0.577, 0.575, and 0.575, respectively.

Table 1: Terms with their weights

Term	Weight
financial	0.583
customer	0.577
team	0.575
pay	0.575
best	0.566
allow	0.564
mental	0.564
safety	0.558
continue	0.554
executive	0.553

These terms collectively draw a better picture of Fortune magazine's 100 Best Workplaces to Work for. For instance, employees in their comments pointed out that they appreciated the fact that they were helped when they were in *financial* need, provided with *financial* incentives and assistance, and that they were all *safe financially* and *mentally*.

Concept Link Diagram

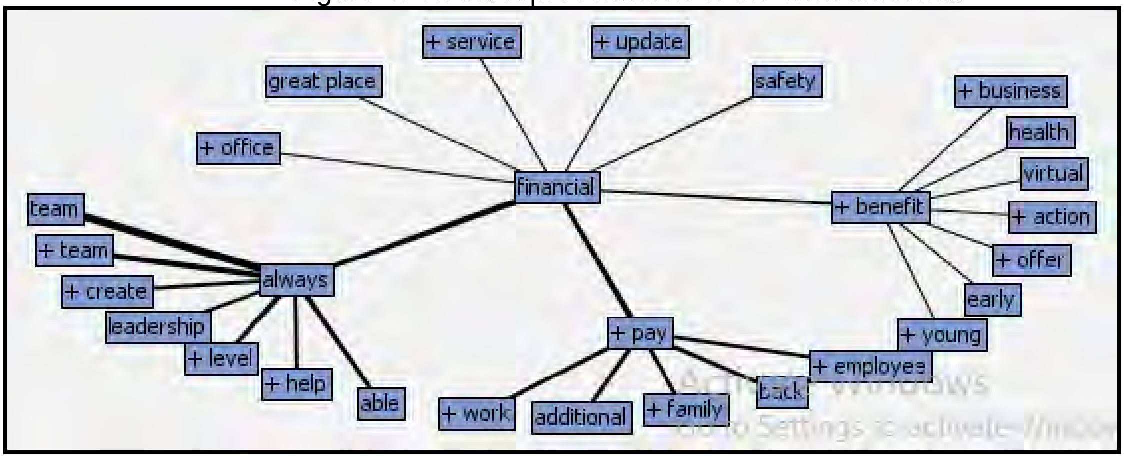
While Table 1 tabulates the most important terms, it does not allow one to see the relationships among the most significant terms found in the comments. Therefore, we generated a concept link diagram to examine what terms are closely associated with one another in the comments. Concept link diagrams are a text mining tool that aids one in visualizing and understanding the relationships among the key terms and words in a document collection.

As tabulated in Table 1, the term "financial" is the most significant term with a weight of 0.583. Figure 4 illustrates the relationship between the term "financial" and the other terms found in the comments made by the employees. Skimming through the comments, various employees speak of how their companies "*helped residents living in their communities with financial assistance*," "*helped those in financial need*," their companies made sure they were all *"safe and healthy physically, mentally, emotionally and financially*," and "*helped parents financially with young children at home.*"

In addition, the term "financial" is strongly associated with the term "always" which in turn has strong relationships with such important terms as "team", "leadership," and "help." Scanning through the comments made by the employees, we see that some employees will "*always be grateful for their leadership during this time,*" their leadership "*is always extremely helpful,*" and that they have "*leadership that is approachable and always willing to listen to their employees' issues or problems.*"

Figure 4 also suggests that employees were grateful for the *financial* and other *benefits offered* by their companies and that their companies would *pay* competitive salaries during the pandemic. It appears, during the pandemic, one of the most important distinguishing characteristics and attributes of the best places to work as ranked by Fortune magazine was the fact that they paid close attention to their employees' financial and mental well-being.

Figure 4: Visual representation of the term financial**.**



Cluster Analysis

Next, we employed the text cluster node, which is used to segment terms into different disjointed sets of themes. The node creates reports on the descriptive terms for those clusters. These terms collectively create “themes” that are paired with the Cluster IDs. The Text Cluster node groups terms and words in a document into clusters based on similarities within them and then provides information on these similar terms. For this analysis, we specified the clustering algorithm as hierarchical clustering to cluster the data into a tree hierarchy, and the maximum number of clusters to 5, with 10 descriptive terms included. These clusters will be made up of 10 descriptive terms from the document.

Table 2 shows the clusters identified by the node. As seen, while cluster 1 is the largest cluster as shown by the frequency of 94 in the output (46%), cluster 3 is the smallest cluster with only 15 terms (0.075%).

Browsing the terms in the clusters, common themes can be identified and the content of the original document that is being examined can be better understood. For instance, cluster 1 contains some key terms such as CEO, stress, support, clear benefits, job, and early. Looking at the comments one can see that employees loved the fact that with the pandemic, their leadership (CEO) communicated *early* and *clearly*, provided care to *support* the health and emotional well-being of employees, offered *leave* to any employee that was feeling uncomfortable coming to work, as well as helped them cope with the emotional *stress.* It appears cluster 1 is all about leadership being clear in both communicating and actions that top priorities were employee wellbeing and safety.

Cluster 2, which is the second largest cluster, speaks of how *management* has given employees the *resources* to be successful during the pandemic, *leadership* making the safety and well-being of *associates* a priority and how they felt valued and secure, *management* understanding of those of them working from home with children, *management* going out of their way to ensure employees have additional budget to set up *home office* space, and how employees were *proud* to be part of such a *great workplace*.

Cluster 4 features such important key terms as covid, work, team, CEO, feel, and leadership. Employees speak of how their *leadership team* has done a superb job of handling the *covid* pandemic, how the *executive team* pays attention to the health and welfare of employees, how they were able to work as a *team*, how they feel their firm care about them and the broader global community, and how *CEO* goes above and beyond to make them *feel* like they are truly part of the company.

Table 2: Cluster Analysis

Cluster ID	Descriptive Terms	Frequency	%
1	early +stress events insurance +support clear benefits +job leave +ceo	94	0.468
2	associates management back +proud resources place +great +office home +keep	55	0.274
3	leave paid customers members health +care working care offered +employee	15	0.075
4	work first covid leadership +know team +feel +team people +ceo	37	0.184

DISCUSSION AND CONCLUSIONS

Text mining allows the analysis of a document such as the one examined in this study to be dissected and understood in a way that is more time efficient than reading the entire document alone. This is helpful to businesses that have large amounts of textual data that they want to uncover the core meaning and magnitude of in a shorter amount of time.

The results reveal common themes in these employers' comments. For instance, the analysis revealed a strong association between positive employee opinions and how senior leaders clearly communicated to their employees the challenges they were facing due to the pandemic. In addition, employees appreciated the fact that they were provided with care to support their health and emotional well-being and helped them cope with emotional stress. It appears leadership played a major role in both communicating and taking action with respect to the fact that their top priorities were employee well-being and safety.

Employees also speak of how management has given them the resources they needed to be successful during the pandemic, leadership making the safety and well-being of associates a priority and how they felt valued and secure, management understanding of those of them working from home with children, management going out of their way to ensure employees have additional budget to set up home office space, and how employees were proud to be part of such a great workplace.

In summary, our analysis revealed strong links between positive employee opinions and open communication between senior leaders and employees, employees' well-being and safety, and how employees were supported financially and emotionally during the pandemic. While employees used some other terms to describe their workplace, their financial and emotional well-being seem to be the most important characteristic of Fortune magazine's "Best Places to Work for."

REFERENCES

Bernardi, R.A., Bosco, S.M., and Vassill, K.M. (2006). Does female representation on boards of directors associate with fortune's “100 best companies to work for” list? *Business & Society*, 45(2), 235-248

Butler, T.D., Armstrong, C., Ellinger, A. and Franke, G. (2016), Employer trustworthiness, worker pride, and camaraderie as a source of competitive advantage: Evidence from great places to work, *Journal of Strategy and Management*, 9(3), 322-343.

Carvalho, A., and Areal, N. (2016). Great places to work: Resilience in times of crisis, *Human Resource Management*, 55(3), 479-498.

Cycyota, C. S., Ferrante, C. J., and Schroeder, J.M. (2016). Corporate social responsibility and employee volunteerism: What do the best companies do? *Business Horizons,* 59(3), 321-329.

DeMaria, K. (2016). Evaluating the internal communications of the triangle's "best places to work, *ELON Journal of Undergraduate Research in Communications*, 7(1).

Feffer, M. (2015). What makes an employer a great place to work?, Retrieved September 13, 2021 from https://www.shrm.org/hr-today/news/hr-magazine/pages/0615-great-places-to-work.aspx.

Filbeck, G. (2001). Mother Jones: Do better places to work imply better places to invest? *Review of Financial Economics*, 10(1), 57-70.

Filbeck, G., and Preece, D. (2003). Fortune’s best 100 companies to work for in America: Do they work for shareholders? *Journal of Business Finance and Accounting*, 30(5-6), 771-797.

Fortune-a (2021). Best companies to work for, Retrieved September 20, 2021, from https://fortune.com/best-companies/2021/search/

Fortune-b (2021). Methodology for 100 best companies to work for (2021), Retrieved September 20, 2021, from https://fortune.com/franchise-list-page/best-companies-2021-methodology/

Henderson, E.M. (2012). Identifying great places to work: a systems framework, Proceedings of the 56th Annual Meeting of the ISSS, San Jose, CA, USA.

Hinkin, T.R., and Tracey, J. B. (2010). What makes it so great?: an analysis of human resources practices among fortune’s best companies to work for, *Cornell Hospitality Quarterly*, 51(2), 158-170.

Nguyen, E. (2014). Text mining and network analysis of digital libraries in R, in *Data Mining Applications with R*. Eds, Zhao, Y., and Cen, Y. Academic Press, Elsevier.

Romero, J. E. (2004). Are the great places to work also great performers? *Academy of Management Perspectives*, 18(2).

Simon, D. H., and DeVaro, J. (2006). Do the best companies to work for provide better customer satisfaction? *Managerial and Decision Economics*, 27(8), 667-683.

Schwartz, T. (2011). The twelve attributes of a truly great place to work, Retrieved September 13, 2021, from https://hbr.org/2011/09/the-twelve-attributes-of-a-tru.html

Vittal, R.S. (2021). Understanding the causes, contingencies and consequences of building best workplaces in the context of 'great places to work in India, *International Journal of Management,* 12(4), 35-40.

DECISION SCIENCES INSTITUTE

What's the Matter with the AI Ethical Issues?

Liqiang Chen
University of Wisconsin – Eau Claire
Email: chenliqi@uwec.edu

ABSTRACT

This research will examine the AI (Artificial Intelligence) induced ethical issues among the college students. We apply Stakeholder Theory and Social Contract Theory to develop a survey questionnaire to explore what ethical issues college students think about using AI in their course work, how AI potentially changes the future job market and what the higher education can do to help students to succeed in their future careers. We plan to collect 200 to 300 samples from multiple universities in U.S. and conduct an ANOVA analysis to answer our research questions.

KEYWORDS: Artificial intelligence, ChatGPT, Stakeholder, Social contract, Survey research, ANOVA

INTRODUCTION

Artificial intelligence (AI) is transforming every aspect of society at the individual, organizational, and societal levels (Brynjolfsson & McAfee, 2017). AI is more than a support tool for making decisions; designing and producing new products and services; and improving personal lives (Brynjolfsson & Mitchell, 2017). It is also changing the employment landscape by replacing existing jobs and creating new jobs (Brynjolfsson & Mitchell, 2017). AI has been one of the most disruptive innovations since the first computer was invented five decades ago, and it is becoming a game-changer in the business community. AI is widely used in almost every industry sector and across every business area (e.g., marketing, management, accounting, finance, supply chain management/operations management).

Modern AI technology is data driven (big data). By taking advantage of fast computing speeds and huge storage capabilities of advanced electronic chips, AI is a mathematical model that makes calculation on a huge amount of dataset. For example, recently ChatGPT has greatly attracted the attention from industries, individuals, and legislatures. ChatGPT is a deep learning neural network model which has 175 billion parameters and consumes 570 GB of training data. Training ChatGPT costs over $12 million.

AI benefits us in many areas, for example, speech recognition, image recognition, DNA testing, auto-drive cars, robots, investment decision support, disease diagnosis and treatment, etc. Recently developed ChatGPT can do even more such as mimicking human conversation, developing and debugging computer programs, writing essays, poetry, song lyrics, composing music, teleplays, stories, answering test questions, producing high-quality images from textual descriptions, etc.

However, AI also brings many issues and potential threats to our daily lives, society, politics, culture, and business community. AI may eliminate some jobs, spread misinformation and disinformation, foster cyberattacks, increase plagiarism and deception, violate privacy, etc. The academia, industries, and policy makers have realized these threats and are working on the regulations and policies to control or minimize them. Since we are still new in these, there are very few legal and ethical actions in place.

LITERATURE REVIEW

Although there are many definitions of AI, one of the most well-known defines AI as the study of how to give features of human intelligence like learning, perception, comprehension, and problem-solving capabilities to a machine (McCarthy et al., 2006). AI consists of computing algorithms and mathematical models implemented in software and hardware such as logical reasoning (e.g., modeling human logical reasoning), knowledge-based systems (e.g., expert inferencing systems), probabilistic approaches (e.g., Bayesian network, fuzzy logic), evolutionary computation, optimization (e.g., genetic algorithm), neural networks, and deep learning. Machine learning is a subset of AI that automatically performs tasks by imitating intelligent human behavior. Deep learning refers to specific machine learning algorithms, primarily an artificial neural network, which has many layers and nodes (neurons) so it can learn and improve its intelligence from large amounts of data. Today, machine learning and deep learning have shown great potential for processing huge datasets and are becoming the most prevalent AI technologies. ChatGPT is a deep learning neural network integrated with a large language model.

Today, there are a huge number of talks, discussions, and research of the good and bad of AI. Dr. Geoffrey Hinton, known as the 'godfather of AI,' along with many industry leaders, scholars and policy makers sound the alarm about the dangers of artificial intelligence and calls for regulating AI development and applications. "AI has spurred anxiety about unemployment, as autonomous systems threaten to replace millions of truck drivers and make Lyft and Uber obsolete. And beyond these larger social and economic considerations, data scientists have real concerns about bias, about ethical implementations of the technology, and about the nature of interactions between AI systems and humans if these systems are to be deployed properly and fairly in even the most mundane applications." (Shaw, 2019) Etzioni (2017) proposed three rules for AI.

1.) An AI system must be subject to the full gamut of laws that apply to its human operator.
2.) An AI system must clearly disclose that it is not human.
3.) An AI system cannot retain or disclose confidential information without explicit approval from the source of that information.

In the IS discipline, IT induced ethical issues have been a research topic, for example, Banerjee et al. (1998) examined individual and situational characteristics of people who ethically/unethically use IT. Recently, Mirbabaie et al. (2022) derived ethical dimensions on AI previous research papers. Our literature review indicates that there is lack of evidence and studies of AI ethics.

RESEARCH QUESTIONS AND THEORETICAL DEVELOPMENT

This research attempts to examine the AI-related ethical issues among college students. Although AI potentially brings many ethical issues, society is not ready to respond to them yet and many concerns have not arrived yet. For example, AI potentially eliminates human workers, but this has not been seen across industries. Recently, ChatGPT not only sends shockwaves across industries but also sparked many concerns in higher education. College students have started using ChatGPT to do their course work. For many college students, ChatGPT is the first AI software they use, and thus it is a good start point to explore the AI's ethical challenge among college students with ChatGPT. Accordingly, our first research question is as follows.

> RQ1: What is the ethical concern and obligation when college students use AI software to do their assignments?

Following up the first research question, college students understand the ethical challenges from AI. Now they may think AI could replace their future jobs if it can do assignments for them. Our second research question is -

> RQ2: Do college students have any concerns about AI disrupting the job market in the future? If yes, what are their major concerns?

If college students are concerned about their future job market, they would like to know what the colleges can do for them. Therefore, our third research question is –

> RQ3: What do college students expect higher education to do for them?

To answer these research questions, we apply two business ethics theories: Stakeholder Theory (Freeman, 1984) and Social Contract Theory (Donaldson & Dunfee, 1994, 1999). Stakeholder Theory argues that a firm should create value for all stakeholders, not just shareholders. Social Contracts Theory posits that the ethical standards that an organization applies depend on the type of norms identified in a given situation; the universal hyper-norms (e.g., core human rights, respect for human dignity, good citizenship) constitute the minimum threshold for ethical behavior and override norms derived locally by firms, industries, relevant communities; different behaviors are recognized to coexist among different situations. Built upon Stakeholder Theory and Social Contract Theory, we will develop a survey questionnaire to answer the three research questions above.

RESEARCH METHODOLOGY

We will conduct a survey in multiple universities in U.S. The study participants are currently enrolled undergraduate students. The questionnaire will be developed and tested in a pilot study first. After a questionnaire is finalized, we will collect data. The expected sample size will be 200 to 300. ANOVA is used to conduct statistical analysis and hypothesis testing.

DISCUSSION AND CONCLUSIONS

In this study, we intend to examine the ethical issues that AI brings up, particularly job elimination. We will conduct an empirical study with a survey questionnaire among college students. This study will answer three research questions regarding using ChatGPT in class assignments, AI replacing human workers, and what higher education can do about them. We realized the college students might not have comprehensive knowledge of whether and how AI may eliminate jobs so the answers to the first research question will lead the research participants to realize the possibility of AI replacing jobs. That is, if AI software can do your assignments, then it can also do your jobs. In addition, we expect the answers will also reveal the college students' thinking of AI induced ethical issues and how higher education would take care of them. In sum, we hope this study will shed light on the AI-related ethical issues, particularly those having high impacts on our lives, society, and business community.

REFERENCES

Banerjee, D., Cronan, T., & Jones, T. (1998). Modeling IT Ethics: A Study in Situational Ethics, *MIS Quarterly*, 22(1), 31-60.

Brynjolfsson, E. & McAfee, A. (2017). The Business of Artificial Intelligence. Harvard Business Review, July 18, 2017. Retrieved from https://hbr.org/cover-story/2017/07/the-business-of-artificial-intelligence.

Brynjolfsson, E. & Mitchell, T. (2017). What Can Machine Learning Do? Workforce Implications. Science, 358(6370), 1530-1534.

Donaldson, T. & Dunfee, T. W. (1994). Toward a Unified Conception of Business Ethics: Integrative Social Contracts Theory. *Academy of Management Review*, 19, 252–284.

Donaldson, T. & Dunfee, T. W. (1999). Ties That Bind: A Social Contracts Approach to Business Ethics. Harvard Business School Press, Cambridge, MA.

Etzioni, O. (2017). How to Regulate Artificial Intelligence. *The New York Times*, Sept. 1, 2017

Freeman, R. E. (1984). Strategic management: A stakeholder approach. Boston: Pitman.

McCarthy, J., Minsky, M. L., Rochester, N. & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AI Magazine, 27(4), 12-14. Retrieved from https://doi.org/10.1609/aimag.v27i4.1904.

Mirbabaie, M., Brendel, A. B., & Hofeditz, L. (2022). Ethics and AI in Information Systems Research. *Communications of the Association for Information Systems*, 50, 726-753. https://doi.org/10.17705/1CAIS.05034

Shaw, J. (2019). Artificial Intelligence and Ethics - Ethics and the dawn of decision-making machines. *Harvard Magazine*, January-February 2019. Retrieved from https://www.harvardmagazine.com/2019/01/artificial-intelligence-limitations.

DECISION SCIENCES INSTITUTE

When Industry 4.0 Meets World Class Manufacturing: Developing a Smart Digital Retrofitting Strategy for Sustainable Manufacturing Operations

Sabine Baumann
Berlin School of Economics and Law, Berlin, Germany
OFFIS Institute for Information Technology, Oldenburg, Germany
sabine.baumann@offis.de

Andrea Bellagarda
Politecnico di Torino, Italy
andrea.bellagarda@polito.it

Claudia Franzè
Politecnico di Torino, Italy
claudia.franze@polito.it

Alessandro Merlino
Politecnico di Torino, Italy
s234899@studenti.polito.it

Danilo Pesce
Politecnico di Torino, Italy
danilo.pesce@polito.it

Jörg Walter
OFFIS Institute for Information Technology, Oldenburg, Germany
joerg.walter@offis.de

ABSTRACT

This paper develops an Industry 4.0 step-by-step implementation strategy drawing from the World Class Manufacturing methodology to achieve sustainable value creation in manufacturing operations. In particular, we focus on the first step of such a strategy – namely Smart Digital Retrofitting – a brownfield approach that integrates new technologies and sensors into legacy systems to achieve Industry 4.0 basic requirements of seamless communication, interoperability among machines, and real-time capability. We characterize this step with the socio-technical system factors, describe its positive outcomes in terms of productivity gains and waste and losses reduction, and discuss its theoretical and practical contributions.

KEYWORDS: Industry 4.0, World Class Manufacturing, Smart digital retrofitting, Operations, Strategy

INTRODUCTION

Industry 4.0 (I4.0) represents a critical paradigm shift for most companies in our world of emergent and continuous changes. Although it has been extensively researched and demonstrated that I4.0 is feasible from a technological point of view, the economic and

management considerations still need to be clarified in light of its implementation in terms of continuous improvement to undertake sustainable development (Ghobakhloo et al., 2021b). The existing literature demonstrates a strong research interest in this area, however, the need for a clear implementation strategy remains a challenge, especially for brownfield sites, which are predominant in the industry. Such a strategy can provide sustainable value creation in manufacturing operations, as decommissioned industrial facilities include assets, machinery, and facilities not designed to be I4.0 compliant (Guerreiro et al., 2018). Specifically, firms need to manage legacy and new technological infrastructures and operations simultaneously to enable an uninterrupted transition toward I4.0 technological solutions; they also need to gradually substitute legacy systems and analog procedures while preventing customer downtime and business disruptions (Cenamor et al., 2019; Park et al., 2020).
Nonetheless the confrontation between legacy arrangements and novel practices can be far from a linear process (Åkesson et al.,2018), thus requiring an adaptation of corporate strategies and objectives to recent advancements in I4.0 technologies (Kretschmer & Khashabi, 2020; Åkesson et al., 2018). Moreover, as I4.0 and the growing availability of data create the need to adjust processes within and across various business lines (Warner & Wäger, 2019), the literature pointed out that processes need to evolve differently, depending on the types of transformed business operations and value chain stages. In particular, marketing, advertising, and supply chain management are early targets of I4.0, as they can elicit fast and sizable gains for firms, whereas operations result more complex and require deeper considerations (Ghobakhloo, 2020; Guenzi & Habel, 2020; Taylor, 2017).
Therefore, in this article, we investigate the following research question: *how should a company develop and implement a structured I4.0 strategy for a brownfield plant for sustainable digitalization of manufacturing operations?*
This paper aims to bridge this gap by demonstrating that the seven-step approach of World Class Manufacturing (WCM) is a suitable digitalization framework for brownfield plants. First, we analyze Smart Digital Retrofitting (SDR) as a sustainable and cost-effective first step to make manufacturing operations I4.0 compliant. SDR is a brownfield approach that integrates new technologies and sensors into existing systems to achieve the basic requirements of I4.0: continuous communication, interoperability between machines, and real-time capabilities (Jaspert et al., 2021). Secondly, we develop an I4.0 implementation strategy following the seven phases of the WCM and characterize each phase with socio-technical system factors. We then describe its positive results in terms of increased productivity and reduced waste and loss and discuss its theoretical and practical contributions.

THEORETICAL BACKGROUND

Industry 4.0: Technology, Strategy, and People

The paradigm shift of the I4.0 has created disruptions across the economic system, altered the value creation paths of the companies, and sowed profound structural changes (Bharadwaj et al., 2013). I4.0 does not concern a single radical invention but encompasses various *"'tech ingredients' that are still evolving into new enabling technologies by convergence and mutual combination"* (Culot et al., 2020, p.5).
Indeed the acquisition and implementation of more advanced technologies severely rely on integrating various traditional technologies (Ghobakhloo et al., 2021a). Advanced technologies refer to the modern technological innovations that have been under development for the past few decades and have now reached a level of maturity that allows them to be available for commercial use in the last decade (i.e., Big Data and Analytics, Cloud Computing, Cyber-

Physical Systems, Digital Twin, Industrial Internet of Things, Virtual and Augmented Reality). Whereas traditional technologies, which are considered prevailing and mature information and operations technologies, play a crucial role in supporting the advanced technologies of I4.0 to perform their intended functions, they have been commercially available for the past few decades and have been commonly used across various industries (i.e., Computer Numerical Control System, Industrial Actuator and Sensors, Industrial Embedded Systems, Intelligent Enterprise Resource Planning, Machine and Process Controller) (Ghobakhloo et al., 2021).
The confluence of such digital technologies triggers the convergence between physical and digital spaces, enhancing the limited connectivity inside the firm and the widespread connectivity along the supply chain (Culot et al., 2020). Given the well-known benefits of such technologies – e.g., productivity and efficiency improvements in the value creation process, advances in flexibility and agility, quality enhancements of value offerings, and cost reductions – companies that have been able to integrate them successfully into their existing business strategies, could stay competitive and promote new growth potentials (Kane et al., 2015; Stief et al., 2016; Hess et al., 2016).
However, transitioning from the traditional industry, with few or no technologies introduced in its processes, to I4.0 practices requires a complete review of operations (Ghobakhloo, 2018). Despite I4.0 paradigms, design principles, and technologies have been extensively examined, more empirical investigations linked with the industrial world must be carried out (Koh et al., 2019). The isolated adoption of technologies generates high investments and expectations but marginal gains in product and service development, thus frustrating managers. On the one hand, adopting a novel Information and Communications Technology before implementing an adequate level of lean practices results in high capital expenditure on wasteful and poorly designed processes. Many companies are still struggling with implementing I4.0 solutions into their operations, as purely technological adoption does not generate the expected results (Tortorella et al., 2019).
On the other hand, I4.0 has been framed as a socio-technical phenomenon, requiring the alignment of technology, strategy, and people (Beier et al., 2020) and, as such, needs to be complemented with adequate digital skills (Firk et al., 2021). This new industrial paradigm changes the roles of human resources and machines in production processes, restructuring the very essence of the workforce concept (Terziyan et al., 2018). Consequently, companies wanting to start their journey toward I4.0 practices must first consider their strategic objectives before implementing any technology (Dalenogare et al., 2018). Therefore, the effectiveness and feasibility of integrating I4.0 into management systems of existing production require more practical insights.

Smart Digital Retrofitting

Brownfield sites, which predominate in the manufacturing industry, are mainly based on heterogeneous technological structures and architectures due to continuous development processes that occurred over the last decades. Indeed, the migration from traditional industry to the I4.0 paradigm presents several issues, as brownfield sites encompass machines and technologies coming from different vendors that are discontinued or no longer updated and were partially replaced by other technologies. As such, replacing old plants with a new-generation plant designed for I4.0 is economically and timely inconvenient and technically challenging, thus opening the doors to the digital retrofit alternative (Ehrlich et al., 2015).
The digital retrofit process involves the addition of new software, hardware, or protocol-related components to existing equipment to expand and update its capabilities and meet new requirements.

This approach is cost-effective – as it avoids the incurring of significant investments required for redesigning or purchasing new equipment – and enables companies to consider a full or partial migration to new technologies while ensuring the continuity of their traditional methods (García et al., 2020). In the context of I4.0, we talk about SDR, which goes beyond traditional retrofitting methods by incorporating I4.0 tools and technology along with the classical retrofitting features (Al-maeeni et al., 2020).
Previous research mainly investigated the technical aspects of retrofitting (Etz et al., 2020; García et al., 2020; Gualtieri et al., 2019), also concerning the improvement of maintenance operations (Cattaneo & Macchi, 2019; Herwan et al., 2019; Hesser & Markert, 2019; Silva et al., 2020), safety operations and assessment (Kummer & Varga, 2019; Wu et al., 2014) and sustainability (Ayani et al., 2018; Lins & Oliveira, 2020; Stock & Seliger, 2016). While we see a growing number of articles with a strong technological approach to this topic, there is still a lack of a socio-technical perspective where the strategic and social aspects are of critical importance (Jaspert et al., 2021). Although the technical changes allow for more efficient processes, they could also negatively impact if those changes are not adequately implemented along with their systemic complementary factors (Pérez et al., 2018).
According to Jaspert et al. (2021), current research on SDR focuses on new technologies and legacy systems but largely disregards transformational aspects regarding the strategy at the organizational level and the interdisciplinary knowledge and skills required for a successful retrofitting process (Jaspert et al., 2021). In fact, *"the complexity, interdisciplinarity, and integrative nature of digital retrofitting make a structured and selectively guided workflow crucial"* (Tantscher & Mayer, 2022, p. 36).

World Class Manufacturing: Principles and Deployment

WCM is a set of concepts, policies, techniques, and principles for operating and managing manufacturing operations that was conceived to deliver world-class performance (Schonberger, 2007). Its purpose is to systematically eliminate waste and losses in manufacturing operations through the rigorous application of standards, methods, and tools. It is based on ten technical and ten managerial pillars (D'Orazio et al., 2020) that need to be developed in parallel to achieve the desired level of excellence in manufacturing operations. Each pillar is mainly linked to a specific area of the production system (such as safety, workplace organization, maintenance, quality, logistics, people competencies, environment, and energy), and through the use of appropriate methods and tools drives its domain to world-class excellence.
For all pillars, WCM defines seven development steps in three phases: reactive, preventive, and proactive (De Felice et al., 2013). These seven steps are designed to guide the implementation of the pillars' activities in a sustainable and integrated way. During the first reactive steps, the purpose of each pillar is to restore basic conditions and standardize the shopfloor and behaviors to ensure they are maintained. In the preventive steps, each pillar focuses on attacking more challenging losses through the use of intermediate methods and tools. Finally, during the proactive steps, each pillar leverages advanced methods and tools to continuously challenge the remaining losses while striving to pursue world-class performance.
Together with the WCM pillars and their seven steps, another fundamental characteristic of the deployment of WCM in manufacturing operations is its model and expansion area approach. Manufacturing processes and flows are first prioritized for each pillar based on the amount of waste and losses they are embedded with. Each pillar is then first implemented in a model area, which is the most critical for the pillar's domain in terms of waste and losses. As the steps or depth in the model area progress, the organization develops knowledge and skills on the correct application of the pillars' methods and tools while also improving an area that, due to its

criticality, maximizes the effectiveness of its activities. After certain depth milestones are achieved, each pillar then expands to other areas, following the waste and losses prioritization previously described. It is essential to underline that the target level of depth and expansion of a given WCM pillar is defined at a strategic level based on the actual benefits expected from the application of the pillar. Therefore, not all areas are covered by a WCM pillar, and not all covered areas are necessarily required to reach the same level of depth as the model area.

THEORETICAL MODEL

Implementing Industry 4.0 with a World Class Manufacturing Framework

As highlighted by Peças et al. (2021), little research was found regarding a holistic application of I4.0's technological concepts towards continuous improvement, which clarifies the potential for improving its effectiveness (Peças et al., 2021). While WCM is mainly based on continuous improvement and loss elimination, Industry 4.0 is based on using all accessible information and data of systems and making decentralized decisions, both involving a global vision and a systemic approach to global profit optimization (Ebrahimi et al., 2019). In this context, the connection between WCM and Industry 4.0 practices paradigm should transform the factory into a fully interconnected environment where decisions can be quickly made based on reliable, accurate, precise, and real-time data (D'Orazio et al., 2020).

Current continuous improvement practices do not have the speed of this new technological standard (Rossini et al., 2019), and there is not much research regarding the holistic application of Industry 4.0 concepts towards continuous improvement, which clarifies the potential for improving its effectiveness (Peças et al., 2021).

However, it is essential to underline that the purpose of this investigation is not to propose a structured approach to integrate lean manufacturing and continuous improvement systems, such as WCM, with Industry 4.0 solutions. The investigation rather focused on developing an innovative framework, based on WCM's seven-step approach, to guide the implementation of Industry 4.0 solutions within manufacturing operations. Within this topic, the purpose of this article is to present the first step of the newly proposed framework, concentrating on SDR as a sustainable and cost-effective first step to making manufacturing operations of a brownfield plant I4.0 compliant.

METHODS

We conducted an exploratory case study of a global automaker company to investigate how a company can develop and implement a structured I4.0 strategy for a brownfield plant for sustainable digitalization of manufacturing operations (Yin, 2017). Such a company recorded a turnover of $100+ bln in 2020, employing over 200.000 people across its 50+ manufacturing plants. In particular, drawing on the principles of WCM and I4.0, we investigated *whether WCM can provide a suitable framework for a company's I4.0 strategy, integrating its digitalization path to achieve sustainable improvements in performance.*

The timeframe for the case study analysis is 5 years, starting from 2017 when the Italian "National Plan Enterprise 4.0", also known as "Piano Nazionale Calenda" – now "Transition Plan 4.0" ("Piano Transizione 4.0") – named after the Minister of Economic Development, was introduced. This plan aimed to promote and foster the adoption of I4.0 technologies, after many years of very minimal attention in industrial policy. It includes measures like "hyper depreciation" and "tax credit for innovation", incentivising investments in technologies, supporting R&D spending, and encouraging training programs to develop the required competencies. The funds

are allocated to machinery and automation solutions, machine maintenance and installation, electrical and electronic equipment, and other investments (MISE, 2017).
We draw on the WCM methodology for the following reasons: (i) it is a philosophy adopted by the company since the early 2000s, therefore, well established at all company levels and practices, especially from the employees' mindset viewpoint; (ii) although it is not a brand-new methodology, its fundamental concept of continuous improvement is more current than ever, especially in the context of the Fourth Industrial Revolution; (iii) it provides a suitable and a systematized framework for better outline and implement the wide concept of I4.0 starting from a model area (manufacturing operations) considered as a pilot. Since 2015, when the I4.0 topic became highly discussed, the company has been reflecting on how to integrate the WCM methodology with the implementation of I4.0 in its manufacturing sites. More specifically, it recognized the risks coming from the implementation of I4.0 solutions without a clear strategy and began reflecting on how to apply WCM's logic and principles to the digital transformation of its industrial sites.

Data Collection

The data were collected during 2022 through multiple data sources to gain a deeper understanding of the dynamics under examination by increasing the information base and diversifying data to reduce biases and triangulate the data to strengthen the reliability and validity of our findings (Patton, 2014).
The primary data collection involved semi-structured face-to-face interviews conducted in Italian and English by at least two researchers following a predetermined topic list. First, we interviewed 11 WCM top and middle managers of the global automaker company; then we interviewed 6 top managers and scientific researchers working in a northern Germany application-oriented research institute with more than 250 employees to ensure a variety of points of view on the same phenomenon under investigation, as well as the granularity necessary for its analysis.
To ensure that experts were knowledgeable and had global visibility on the phenomenon, the expert panel was chosen according to expertise and long-time practical experience on both the topic of WCM and I4.0. Moreover, their heterogeneous background provides an informed opinion on the issues in its different facets.
The interview method was chosen to gather both retrospective and real-time narratives from individuals who have experienced the phenomenon of theoretical interest (Gioia et al., 2013). The interviews lasted from 1 to 3 hours each, were tape-recorded integrally for an overall total of 48,5 hours of interviews, transcribed verbatim, and then translated into English with the help of an independent native speaker, for a total of 330 pages of transcript. An overview of the interviews is shown in Table 1 and Table 2

Table 1: Interviewees of the global automaker company

Interviewees	*In the company since*	*Number*	*Duration (h)*	*Period*
WCM Director Europe and WCM Master Auditor	30+ years	2	2h 2h	15.03.22 29.03.22
WCM Global Autonomous and Professional Maintenance Leader	15+	4	2h 2h 1.5h 1h	11.04.22 15.04.22 27.04.22 04.05.22

WCM Global Cost Deployment Leader	20+	2	1.5h 1.5h	22.04.22 27.04.22
WCM Senior Specialist	20+	2	2h 2h	04.05.22 01.06.22
WCM Global Audit System	20+	1	0.5	01.06.22
International Manufacturing Engineering Leader	30+	3	3h 3h 3h	06.06.22 13.06.22 14.06.22
Global Head of WCM Academy	20+	1	3h	13.07.22
WCM Logistics Pillar Leader	20+	1	3h	14.07.22
WCM Senior Specialist	20+	1	3h	14.07.22
ICT and WCM Manager	20+	1	2h	15.07.22
WCM Innovation and Industry 4.0 Leader	20+	3	3h 3h 3h	16.03.22 26.04.22 19.07.22
WCM Powertrain Maintenance Specialist	30+	1	2h	15.07.22

Table 2: Interviewees of the research institute

Interviewees	*In the research institute since*	*Number*	*Duration (h)*	*Period*
Director R&D Division Manufacturing	25+	1	1h	10.10.22
Group Manager - Distributed Computing and Communication	10+	3	2h 1h 2h	10.10.22 18.10.22 16.11.22
Scientific researcher - Distributed Computing and Communication	5+	2	2h 1h	19.10.22 6.12.22
Scientific Researcher - Manufacturing Operations Management	5+	1	1h	5.10.22
Scientific Researcher - Sustainable manufacturing systems	5+	2	2h	19.10.22
Group Manager – Manufacturing Operations Management	10+	1	1h	18.10.22

We also gathered qualitative and quantitative data from secondary sources, that is, from publicly available information and internal documents, for a total of 350 pages to complement primary data from interviews (Eisenhardt & Graebner, 2007; Yin, 2017). Table 3 describes the secondary data sources.

Table 3: Secondary data collection	
Description of the data source and year	*Evidence (n. of pages)*
Publicly available information	
White papers	40
Archival documents	90
Reports	55
Databases	-
Website	-
Internal documents	
Financial statements of the global automaker company from 2014 to 2021	75
Company WCM Audits	55
Other technical documents and material provided by the informants	35
Total number of evidence	*350*

Data Analysis

We assessed secondary data sources to ensure their overall suitability for our research questions and objectives. We paid particular attention to the measurement validity and coverage of the data. Then we evaluated their particular suitability, including reliability for our research (Saunders et al., 2012). The analysis of secondary data was the starting point for creating a first understanding of the case company to identify and aggregate an initial I4.0 implementation pattern. Secondary data was then triangulated with other sources to help us refine and strengthen our emerging interpretations (Yin, 2016).

Quantitative secondary data were useful for tracing the company's growth path in terms of size and turnover, considering the investments made in digitalization. Following the prescriptions for case studies, we analyzed our data through multiple rounds of coding (Gioia et al., 2013).

Coding and measurements were implemented to reduce the potential of confirmation bias affecting the results and growing descriptive and theoretical validity (Strauss & Corbin, 2014).

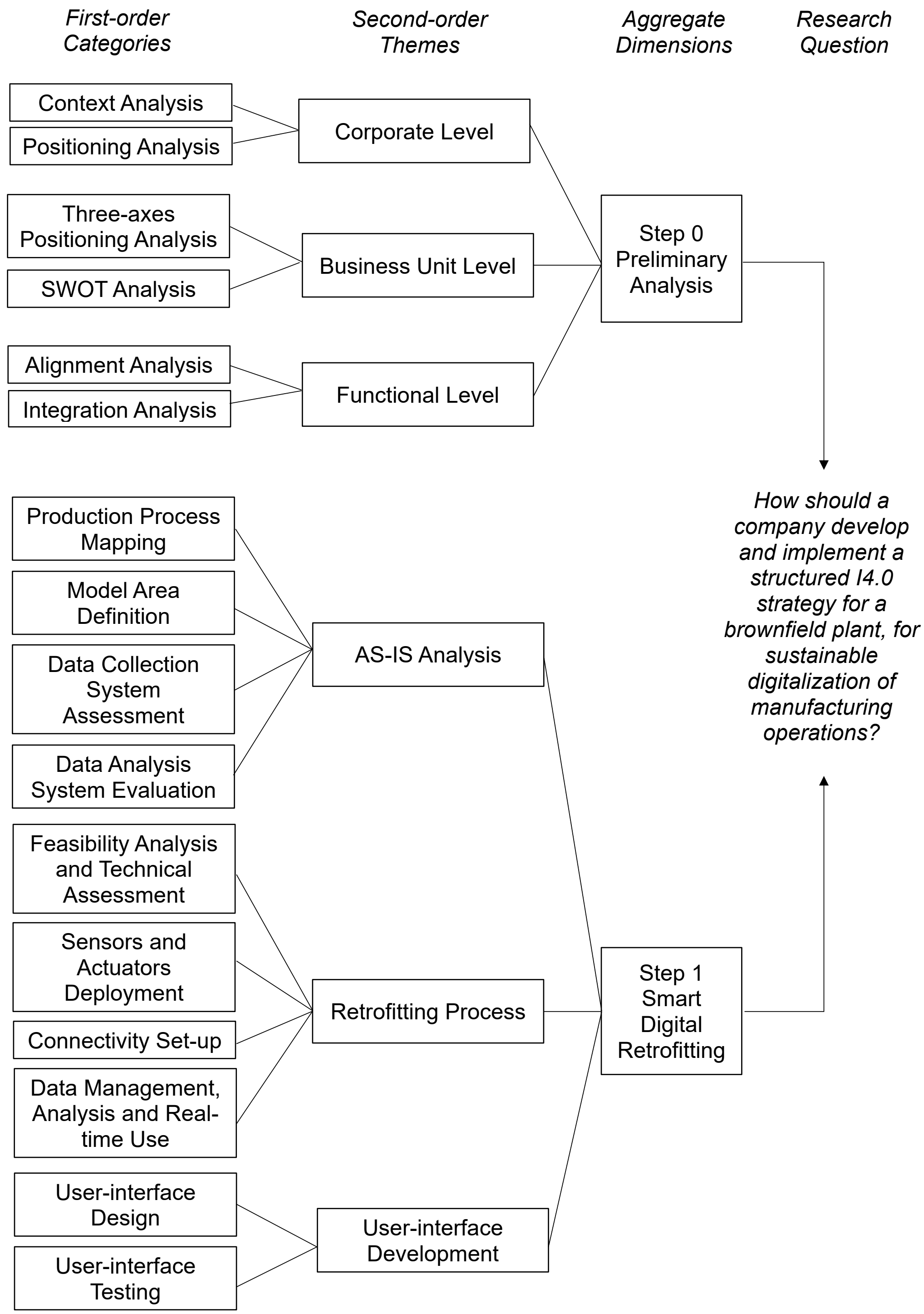
We coded the SDR implementation process of the company according to the WCM principles and practices and searched the data for references to the technological, strategic, and social SDR complementary factors. To verify the reliability of our coding, we individually checked the previously coded text to ensure the internal consistency of the emerging coding structure (Miles & Huberman, 1994), and we performed periodic comparisons between us (Locke, 2001). Through iterations of feedback we have been able to reach a convergence of the experts' opinions.

The inductive analysis of the primary and secondary sources allows concepts and relationships to emerge from the data to identify how a company can develop and implement a structured I4.0 strategy for a brownfield plant for sustainable digitalization of manufacturing operations. We iterated back and forth between the empirical data and emergent theory to identify themes and overarching dimensions to develop an empirically grounded framework.

Then, we disentangled the I4.0 implementation according to the seven steps of WCM, and we characterized each step with the socio-technical system factors. Data were coded into first-order categories, and these were then clustered into second-order themes, which converged into aggregate dimensions (Gioia et al., 2013). We constantly updated and refined our emerging framework based on evidence gathered in subsequent interviews (Burawoy et al., 1991).

The results of the coding procedure are presented in Figure 1 and will be described more extensively in the Results section.

Figure 1: Results of the coding procedure



In the final step of our analysis, a pattern-matching technique was used to compare the predicted theoretical pattern with the observed empirical pattern (Yin, 2017).

RESULTS

Drawing from the WCM methodology and our data analysis, we derived an I4.0 step-by-step implementation strategy. We focus particularly on the first step, identifying SDR as a suitable approach to make manufacturing operations I4.0 compliant and characterizing it with the socio-technical system factors – (i) a sequence of strategic and technical activities required to move to the next step; (ii) the main criticalities to be solved to build a solid foundation/path; (iii) the specific skills required by the company's roles.

Step 0: Preliminary Analysis

It is essential to lay the foundations for an informed and successful long-term strategy.
Its (i) strategic and technical activities are carried out at three different organizational levels: corporate, business unit, and functional.
At the corporate level, the goal is to achieve an overall excellent condition to maximize the profits obtained globally from all the different business units and functions of the company. Two types of analysis should be carried out: context and positioning analysis. For the context analysis, a useful tool is represented by the PESTEL analysis of the surrounding world's Political, Economic, Social, Technological, Environmental, and Legal situation. It is used to determine the external factors and understand the context in which the company operates to properly define goals, priorities, and constraints from the external context. For the positioning analysis, a three-axis tool (level of service, inventory level, operative expenses) is suitable for developing the business strategy that is intended to be adopted. Such a strategy is driven by factors such as market competitiveness and customer response to the company's choices. Each company can choose its positioning independently, considering the type of product or service it offers, usually through a benchmark with other competitors.
The three-axis tool can also be used in the business unit level analysis. In this case, every business unit refers to different targets and has different needs and goals, mainly in terms of level of service. This positioning is not independent from the corporate one since the board defines the company's overall objectives, which are reached through the results of each unit. Then a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis aims to determine the business unit's competitive positioning by considering and evaluating internal and external elements.
At the functional level – in this case, the manufacturing function is considered – the proposed I4.0 step-by-step implementation strategy can be applied. It is necessary to keep in mind that Manufacturing itself is one of the business functions, so superimposed inputs and constraints must be respected (alignment analysis). Coordination between this function and all the other company functions or departments is fundamental to reaching the global excellence of the company (integration analysis). This means that some specific solutions designed to solve highly specialized aspects of production cannot be implemented at the corporate level because the optimal situation for the company – which is achieved by harmonizing and managing all functions as a whole – does not always correspond to the optimal situation in a single department.
The (ii) criticalities to be solved in this Step 0 are mainly to identify the strategic goals of the retrofitting in terms of I4.0 targets, such as improving working conditions, better quality processes, better communication, collaboration, improved productivity, improved efficiency,

flexibility, agility, or reduced costs. The scope is to evaluate the current status of digitalization through an assessment of the company, an evaluation of current and future resources, and a study of the possible incentives. The outcomes are an Assessment Score and Competence Map, a Benefits and Costs estimation, and a Digital Strategic Plan (including objectives, KPIs, schedule and milestones, and incentives plan).
Finally, the (iii) skills required for top and middle management are mainly analytical, managerial, strategic, and technical.

Step 1: Smart Digital Retrofitting

This step concerns the following three phases: (1) as-is analysis, (2) retrofitting process, and (3) user-interface development.
In the as-is analysis phase (1), a maturity assessment of the industrial process in terms of data collection and management is carried out.
First, the production process needs to be mapped by developing an accurate description of the plant and process to analyze the plant's main functions and identify bottlenecks and critical areas to assess how lean the process is. Then the variables (e.g., temperature, pressure, flow rate) that come into play in the process, the variables already monitored, and those that must be controlled are identified. Significant issues in the process must be solved within this step since moving to the next phase will only be possible if the production system is lean. Therefore, tools from the existing pillars of WCM can be applied to determine how to reduce waste and losses in the process. For instance, valuable tools from the Cost Deployment pillar are Matrix A – which shows where the inefficiencies are in the process and their entity - and Matrix B - which correlates the losses and their root causes in order to act on them directly.
Second, a model area should be defined. Following the WCM approach, the improvement actions are first taken within a Model Area – that generally is the most critical one (e.g., the machine with the highest number of failures) – and then the measures are brought to an Expansion Area, up to be implemented in the whole plant. Therefore, the definition of the Model Area is crucial to get results quickly and develop the experience and know-how necessary to expand the improvement actions throughout the process.
Third, the data collection system should be assessed by investigating which and how the data is collected and distinguishing between basic and advanced data. Basic data focuses on high-level production performance, such as production, quality, and costs/resource consumption. This data must be collected at the end of the production process and specific critical points (bottlenecks), as they are always necessary for an overview of the production. Activities auxiliary to production can also be highlighted when determining critical data to control the operations. Advanced data is more detailed data (e.g., OEE, MTBF, MTTR) for in-depth monitoring and analysis of the performance of a specific and critical portion of the process. It is unnecessary to have them run the process in a traditional way. As the lean manufacturing system is implemented, these data become an essential source of information to make improvements and to take the production system and all the auxiliary activities under control. Identifying which type of data is necessary is a critical step that will affect the goal definition and the future benefits of SDR.
Finally, the data analysis system should be evaluated. Since the data collection system has been analyzed, it is necessary to investigate how data are exploited. If they are used to monitor and improve the production system, then the system that the company is already using to manage data can be enhanced such that the digitalization process can start. On the contrary, if the company uses no data or the system is not mature enough, a data management system

must be implemented to exploit data coming from production to make the system lean and reduce wastes and losses with the already available data.
In summary, what is required by this step in digital terms is to understand if there are sensors along the line, if they are working and collecting data, and if there is a usage of the collected data. The output of this analysis provides essential information about the process and the capabilities of the company to use and manage data properly.
Two restrictions must be observed to move on to the next step. If the system is not lean, the company cannot move to the next step to avoid the digitalization of waste and losses, which would lead to inefficient use of the resources. Whereas, if the company does not have a system made by software and good practices to manage data, the next step cannot be undertaken because the company is not mature enough to move to the digital approach.
These constraints aim to avoid the arise of digital wastes or inefficiencies in the data management system, which would make the effort in digitalization vain and the investment unprofitable. Moreover, the foundations for a lean manufacturing system are laid since the prerequisites in terms of data collection and management that are defined in this step are also the basic information required to implement a lean manufacturing system, which is fundamental for a manufacturing company that wants to shift towards the I4.0 paradigm.
The (ii) criticalities of the as-is analysis phase to be solved are mainly the clear distinction between basic data and advanced data since it will affect the Smart Digital Retrofitting goals and future benefits.
The second phase of Step 1 is the retrofitting process (2), a strategic technical guide for first-time quality implementation and predisposition for I4.0 scalability. As the vision concerns start from a model area towards implementing I4.0 in the whole plant, a future-oriented mindset is necessary. This phase is also based on four (i) strategic and technical activities.
First, feasibility analysis and a technical assessment are carried out to evaluate all the necessary technical interventions at the hardware and software levels and all related costs (design, acquisition, installation, operation, maintenance, and disposal); and development and installation time (machine downtime). The necessity and cost (time and money) of establishing a comprehensive catalog of existing, employed technologies, communication standards, and/or I/O interfaces are functional for the following activities. In fact, SDR is challenging without this information, as it can be astonishingly time-consuming to find out details about a 30-year-old machine, for instance (that otherwise works well). Such information should be collected systematically and in a structured way; otherwise, when collecting it on-demand, the result is an unpredictable critical path. It also might happen that some information is irretrievably lost; thus, it is necessary to know that to plan an alternative better. Afterward, it is essential to find suppliers for Internet of Things (IoT) devices/hardware components that can interface with those technologies/standards. Such a process might be either expensive and/or time-consuming.
Second, sensors and actuators are deployed by defining the strategic points to record further physical values/variables such as vibrations, temperatures, or cycle times. Upgrading the infrastructure of the legacy machines requires retrofitting smart sensors, actuators, and IoT devices in order to establish a starting point for integrating physical and virtual worlds. These strategic points should be made usable to guarantee data extraction from machines. Therefore, controllers such as microcontrollers, PLCs, or computers are introduced in this step and referred to as IoT nodes. The introduction of IoT nodes ensures the data extraction from the sensors and connects the machine to the internet to facilitate the next steps. Then, a secure gateway between the physical and cyber layers should be established by introducing a middle layer called Industrial Internet of Thing (IIoT) gateway, which is further connected to IIoT middleware. For this purpose, a standardized communication protocol is required (OPC UA): this protocol ensures accurate and swift data transfer between the machine and the IIoT middleware. Finally,

each retrofit device itself should be required to support OPC UA to keep the impact (technical risk, efficiency loss, monetary cost) of legacy technology mismatch as localized and independent as possible.
The last activity of the retrofitting process concerns data management, analysis, and real-time use. As basic/advanced data are digitalized for Smart Data Collection, the collected data is stored in databases, sent for analytical purposes using a web-based software or local edge-computing device, and used in real-time.
The (ii) criticalities to be solved are mainly to create a future-oriented mindset, which is necessary to start from the model area and towards the whole plant.
Finally, the third phase of Step 1 is the user-interface development (3) to make the developed applications easy to use for the operators. Although operators should have experienced at least some training and are known, the interface must be designed to operate by very heterogeneous user groups with little to no training in varying contexts and environments and with only a little knowledge about the users themselves. Its (i) strategic and technical activities are only the development of such a user interface. The (ii) criticalities to be solved are mainly the development of a user interface that is easy to use for the operators.
In terms of skills (iii) of the whole SDR step: (a) operational roles no longer manually enter data but read, understand, and verify the correctness of such data. To do so, basic data analytics skills complement their legacy skills; (b) middle management and process technicians can actually exploit the data collected by strong analytical and data-driven decision skills, in fact, digital technologies require employees to depend more heavily on their analytical skills to solve increasingly complex business problems; (c) top management should be aware of the potential of data and lead the employees towards its exploitation.

<u>Preliminary benefits of an Industry 4.0 step-by-step implementation strategy</u>

The following quantitative results have been provided by the company thanks to the processing performed by its customized and proprietary data collection system, which is specific for identifying waste and losses. In addition to gathering a significant amount of data, this system is also capable of integrating various types of data and cleaning them in order to make them usable for analyzing various KPIs.
We found that the I4.0 implementation in a brownfield setting of automotive manufacturing facilities, together with the application of a lean methodology such as WCM, improves manufacturing data collection accuracy (+85%) and speed (+60%). This reduces the time required to complete improvement projects (-20%), and increases the accuracy of the expected savings (+15%), enabling greater productivity (+20%) while lowering maintenance costs (-23%), downtime (-35%) and quality costs (-25%).
We also found on the one hand that the benefits of WCM on I4.0 concern avoiding automation and digitalization of wastes and knowing what to measure and what to control; on the other hand that the benefits of I4.0 on WCM include the elimination of shortcuts wastes and the use of a reliable source of data to measure improvements.

DISCUSSION AND CONCLUSIONS

To investigate how a company can develop and implement a structured I4.0 strategy for a brownfield plant for sustainable digitalization of manufacturing operations, we conducted an exploratory case study of a global automaker company (Yin, 2017) developing an I4.0 implementation strategy drawing on the principles of WCM.

We propose a preliminary analysis step and a smart digital retrofitting step and characterize each step with the socio-technical system factors. I4.0 requires significant investments from companies, and it is, therefore, necessary to make strategic and careful choices to achieve the planned results – thus making the investment profitable and pursuing sustainable value creation as an outcome of the undertaken actions. However, our results highlight the importance of a Lean Manufacturing system, considered a prerequisite for implementing a successful digital production system. We here discuss three main points.

First, the smart digitalization of existing production processes allows identifying waste and losses more quickly and accurately, guiding plant management to understand the improvements which must be implemented from either a technical or social viewpoint to enhance production performance, for example extending the life cycle of machinery and equipment in a way that is feasible, time-saving, and requires comparatively low investments concerning the replacement (García et al., 2020; Jaspert et al., 2021). I4.0 enables novel forms of organization of manufacturing processes. The distributed model of production can help realize I4.0 technologies' sustainability potential by creating favorable conditions for the development of resource-efficient and sustainable manufacturing (Kumar et al., 2020). Distributed manufacturing's contribution to sustainability is realized through shorter lead times; an emphasis on production to order, rather than make-to-stock; and a modular production process (Srai et al., 2020).

Second, we argue that the SDR should be supported, developed, and implemented by a dedicated team led by a Digital Transformation Officer, a hybrid figure with a mix of managerial and technical skills (Firk et al., 2021). In fact, on the one hand, our findings highlight the need for analytical, managerial, strategic, and technical skills, specifically the complementarity between basic data analytics skills and legacy skills, followed by data-driven decision skills. On the other hand, the need for creating a distinct and central role to handle the digital transformation successfully is fundamental as digital transformation is associated with a shift in corporate strategy toward differential value creation based on digital resources and skills, and it transcends functional and divisional boundaries thus requiring central coordination.

Third, as stated by Terra et al. (2021), we are aware of the difficulties in integrating I4.0 and WCM due to the rapid expansion of production systems and the increasingly complex data monitoring (Terra et al., 2021). Despite the positive outcomes of linking a company's strategy to Industry 4.0 practices and the World-Class Manufacturing method, some challenges still remain, such as the costs associated with the use of technologies, the lack of knowledge of the applied methods and tools, the lack of trained and qualified human resources, and the resistance of people to the use and application of the newly adopted practices. All these factors, mainly related to the lack of staff and finance, might led to losing interest into projects in early stage.

Our study contributes to theory as it refines and enhances the WCM approach – by integrating the I4.0 concept – and systematizes the I4.0 concept – by disentangling its implementation into seven incremental steps with a socio-technical system perspective.

The practical contribution for managers is the definition of a model to guide investments in I4.0 technologies for an optimal implementation of the I4.0 socio-technical system. Specifically, smart digitalization of existing production processes allows managers to identify waste and losses more quickly and accurately, helping them understand the improvements which must be implemented from either a technical or social viewpoint to enhance production performance.

The contribution for policymakers is the definition of industrial and labor policies to enable firms to accumulate competencies and achieve competitiveness improvements.

As this study is part of a bigger research project, future research will focus on the other steps of the I4.0 implementation strategy. Simultaneously, further research is needed in the automotive sector by analyzing other companies and with multiple case studies in different industrial sectors to generalize our findings and study the effects at a macroeconomic level.

REFERENCES

Al-maeeni, S.S., Kuhnhen, C., Engel, B., & Schiller, M. (2020). Smart retrofitting of machine tools in the context of industry 4.0. *Procedia CIRP, 88*, 369-374.

Ayani, M., Ganebäck, M., & Ng, A.H.C. (2018). Digital Twin: Applying emulation for machine reconditioning. *Procedia CIRP*, *72*, 243–248.
Åkesson, M., Sorensen, C., & Eriksson, C. I. (2018). Ambidexterity under digitalization: A tale of two decades of new media at a Swedish newspaper. *Scandinavian Journal of Management, 34*(3), 276–288

Beier, G., Ullrich, A., Niehoff, S., Reißig, M. & Habich, M. (2020), Industry 4.0: How it is defined from a sociotechnical perspective and how much sustainability it includes – A literature review, *Journal of Cleaner Production*, *259*, 120856.

Bharadwaj, A., El Sawy, O., Pavlou, P. & Venkatraman, N. (2013). Digital Business Strategy: Toward a Next Generation of Insights. *MIS Quarterly*, *37*(2), 471–482.

Burawoy, M., Burton, A., Ferguson, A., & Fox, K. (1991). *Ethnography unbound: Power and Resistance in the Modern Metropolis.* University of California Press.

Cattaneo, L. & Macchi, M. (2019). A Digital Twin Proof of Concept to Support Machine Prognostics with Low Availability of Run-To-Failure Data. *IFAC-PapersOnLine*, *52*, 37–42.

Cenamor, J., Parida, V., & Wincent, J. (2019). How entrepreneurial SMEs compete through digital platforms: The roles of digital platform capability, network capability and ambidexterity. *Journal of Business Research, 100*, 196–206.

Culot, G., Nassimbeni, G., Orzes, G. & Sartor, M. (2020). Behind the definition of Industry 4.0: Analysis and open questions. *International Journal of Production Economics*, *226*, 107617.

D'Orazio, L., Messina, R. & Schiraldi, M.M. (2020). Industry 4.0 and World Class Manufacturing Integration: 100 Technologies for a WCM-I4.0 Matrix. *Applied Sciences*, *10*(14), 4942.

Dalenogare, L.S., Benitez, G.B., Ayala, N.F. & Frank, A.G. (2018). The expected contribution of Industry 4.0 technologies for industrial performance. *International Journal of Production Economics*, *204*, 383–394.

De Felice, F., Petrillo, A. & Monfreda, S. (2013). Improving Operations Performance with World Class Manufacturing Technique: A Case in Automotive Industry. In: Schiraldi M. (Ed.), *Operations Management*. InTech

Ebrahimi, M., Baboli, A. & Rother, E. (2019). The evolution of world class manufacturing toward Industry 4.0: A case study in the automotive industry. *IFAC-PapersOnLine*, *52*(10), 188–194.

Ehrlich, M., Wisniewski, L., & Jasperneite, J. (2015). Usage of Retrofitting for Migration of Industrial Production Lines to Industry 4.0. In: *KommA 2015 - Jahreskolloquium Kommunikation in der Automation*, 6, Magdeburg, Germany.

Eisenhardt, K.M. & Graebner, M.E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32.

Etz, D., Brantner, H. & Kastner, W. (2020). Smart Manufacturing Retrofit for Brownfield Systems. *Procedia Manufacturing, 42*, 327–332.

Firk, S., Hanelt, A., Oehmichen, J. & Wolff, M. (2021). Chief Digital Officers: An Analysis of the Presence of a Centralized Digital Transformation Role. *Journal of Management Studies*, *58*(7), 1800–1831.

García, J.I., Cano, R.E. & Contreras, J.D. (2020). Digital retrofit: A first step toward the adoption of Industry 4.0 to the manufacturing systems of small and medium-sized enterprises. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *234*(8), 1156–1169.

Ghobakhloo, M. (2018). The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of Manufacturing Technology Management*, *29*(6), 910–936.

Ghobakhloo, M. (2020). Determinants of information and digital technology implementation for smart manufacturing. *International Journal of Production Research, 58*(8), 2384–2405

Ghobakhloo, M., Fathi, M., Iranmanesh, M., Maroufkhani, P. & Morales, M.E. (2021a). Industry 4.0 ten years on: A bibliometric and systematic review of concepts, sustainability value drivers, and success determinants. *Journal of Cleaner Production*, *302*, 127052.

Ghobakhloo, M., Iranmanesh, M., Grybauskas, A., Vilkas, M. & Petraitė, M. (2021b),. Industry 4.0, innovation, and sustainable development: A systematic review and a roadmap to sustainable innovation. *Business Strategy and the Environment*, *30*(8), 4237–4257.

Gioia, D.A., Corley, K.G. & Hamilton, A.L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15–31.

Gualtieri, L., Rauch, E., Vidoni, R. & Matt, D.T. (2019). An evaluation methodology for the conversion of manual assembly systems into human-robot collaborative workcells. *Procedia Manufacturing, 38*, 358–366.

Guenzi, P., & Habel, J. (2020). Mastering the digital transformation of sales. *CaliforniaManagement Review, 62*(4), 57–85

Guerreiro, B. V., Lins, R.G., Sun, J. & Schmitt, R. (2018). Definition of Smart Retrofitting: First Steps for a Company to Deploy Aspects of Industry 4.0. In Hamrol A., Ciszak, O., Legutko S. & Jurczyk M. (Eds.), *Advances in Manufacturing. Lecture Notes in Mechanical Engineering*, Springer, 161–170.

Herwan, J., Kano, S., Ryabov, O., Sawada, H., Kasashima, N. & Misaka, T. (2019),. Retrofitting old CNC turning with an accelerometer at a remote location towards Industry 4.0. *Manufacturing Letters, 21*, 56–59.

Hess, T., Benlian, A., Matt, C. & Wiesböck, F. (2016). Options for formulating a digital transformation strategy. *MIS Quarterly Executive*, *15*(2), 103–119.

Hesser, D.F. & Markert, B. (2019). Tool wear monitoring of a retrofitted CNC milling machine using artificial neural networks. *Manufacturing Letters, 19*, 1–4.

Jaspert, D., Ebel, M., Eckhardt, A. & Poeppelbuss, J. (2021). Smart retrofitting in manufacturing: A systematic review. *Journal of Cleaner Production*, *312*, 127555.

Kane, G.C., Palmer, D., Philips Nguyen, A., Kiron, D. & Buckley, N. (2015). Strategy, Not Technology, Drives Digital Transformation. *MIT Sloan Management Review & Deloitte*, *14*, 1–25.

Koh, L., Orzes, G. & Jia, F. (2019). The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management. *International Journal of Operations & Production Management*, *39*(6/7/8), 817–828.

Kretschmer, T., & Khashabi, P. (2020). Digital transformation and organization design: An integrated approach. *California Management Review, 62*(4), 86–104.

Kumar, M., Tsolakis, N., Agarwal, A., & Srai, J. S. (2020). Developing distributed manufacturing strategies from the perspective of a product-process matrix. *International Journal of Production Economics, 219*(1), 1–17.

Kummer, A. & Varga, T. (2019). Process simulator assisted framework to support process safety analysis. *Journal of Loss Prevention in the Process Industries, 58*, 22–29.

Lins, T. & Oliveira, R.A.R. (2020). Cyber-physical production systems retrofitting in context of Industry 4.0. *Computers & Industrial Engineering*, *139*, 106193.

Locke, K. (2001). *Grounded Theory in Management Research*, SAGE Publications.

Miles, M.B. & Huberman, A. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA, SAGE Publications.

MISE. (2017). The Diffusion Of Industry 4.0 Companies And Policies: 2017 Evidence.

Park, Y., Pavlou, P. A., & Saraf, N. (2020). Configurations for achieving organizational ambidexterity with digitization. *Information Systems Research, 31*(4), 1376–1397.

Patton, M.Q. (2014). *Qualitative evaluation and research methods*. Fourth Edi, SAGE Publications.

Peças, P., Encarnação, J., Gambôa, M., Sampayo, M. & Jorge, D. (2021). PDCA 4.0: A New Conceptual Approach for Continuous Improvement in the Industry 4.0 Paradigm. *Applied Sciences*, *11*(16), 7671.

Pérez, J.D., Buitrón, R.E., & Melo, J.I. (2018). Methodology for the Retrofitting of Manufacturing Resources for Migration of SME Towards Industry 4.0. In: Florez, H., Diaz, C., Chavarriaga, J. (Eds.), *Applied Informatics. ICAI 2018. Communications in Computer and Information Science*, 942. Springer, Cham.

Rossini, M., Costa, F., Tortorella, G.L. & Portioli-Staudacher, A. (2019). The interrelation between Industry 4.0 and lean production: an empirical study on European manufacturers. *The International Journal of Advanced Manufacturing Technology*, *102*(9–12), 3963–3976.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research Methods for Business Students* (6th ed.). Pearson.

Schonberger, R.J. (2007). Japanese production management: An evolution—With mixed success. *Journal of Operations Management*, *25*(2), 403–419.

Silva, M.F., Luís Lima, J., Reis, L.P., Sanfeliu, A. & Tardioli, D. (2020). *Advances in Intelligent Systems and Computing*. Springer International Publishing: Cham, Switzerland.

Srai, J. S., Graham, G., Hennelly, P., Phillips, W., Kapletia, D., & Lorentz, H. (2020). Distributed manufacturing: A new form of localised production? *International Journal of Operations & Production Management, 40*(6), 697–727.

Stief, S.E., Edhoff, A.T. & Voeth, M. (2016). Transform to succeed: an empirical analysis of digital transformation in firms. *International Journal of Economics and Management Engineering*, *10*(6), 1833–1842.

Stock, T. & Seliger, G. (2016). Opportunities of Sustainable Manufacturing in Industry 4.0. *Procedia CIRP, 40*, 536–541.

Strauss, A. L., & Corbin, J. (2014). *Basics of Qualitative Research* (Fourth Edi). SAGE Publications.

Tantscher, D. & Mayer, B. (2022). Digital Retrofitting of legacy machines: A holistic procedure model for industrial companies. *CIRP Journal of Manufacturing Science and Technology*, *36*, 35-44.

Taylor, A. (2017). Social work and digitalisation: Bridging the knowledge gaps. *Social Work Education, 36*(8), 869–879.

Terra, J.D.R., Berssaneti, F.T. & Quintanilha, J.A. (2021). Challenges and barriers to connecting World Class Manufacturing and continuous improvement processes to Industry 4.0 paradigms. *Engineering Management in Production and Services*, *13*(4), 115–130.

Terziyan, V., Gryshko, S. & Golovianko, M. (2018). Patented intelligence: Cloning human decision models for Industry 4.0. *Journal of Manufacturing Systems*, *48*, 204–217.

Tortorella, G.L., Giglio, R. & van Dun, D.H. (2019). Industry 4.0 adoption as a moderator of the impact of lean production practices on operational performance improvement. *International Journal of Operations & Production Management*, *39*(6/7/8), 860–886.

Warner, K. S. R., & Wäger, M. (2019). Building dynamic capabilities for digital transformation: An ongoing process of strategic renewal. *Long Range Planning, 52*(3), 326–349.

Wu, J., Zhang, L., Hu, J., Lind, M., Zhang, X., Jørgensen, S.B., Sin, G. & Jensen, N. (2014). An integrated qualitative and quantitative modeling framework for computer-assisted HAZOP studies. *AIChE Journal, 60*, 4150–4173.

Yin, R. (2017). *Case study research and applications: design and methods*. Sixth Edit, Los Angeles, SAGE Publications.

DECISION SCIENCES INSTITUTE
When Is Genetic Engineering Going Too Far?

Aurore J. Kamssu
Tennessee State University
Email: akamssu@tnstate.edu

Eric Philip
Tennessee State University
Email: ephilip@my.tnstate.edu

Mustafa Alsadeq
Tennessee State University
Email: malsadeq@my.tnstate.edu

Ali Nasib
Tennessee State University
Email: anasib@my.tnstate.edu

ABSTRACT

Genetic Engineering and Genetic Technology has been advancing exponentially through time and has a myriad of benefits in health, economics, the environment, and other fields. However, human error and malevolent purposes have led to anthropogenic malpractice of these genetic means. A primary example of these malpractices is using genetic engineering in botanic, however, creating potential diseases which can spread to other plants and cause the death of ecosystems. As humans, our understanding of these tools is rapidly expanding however we still have a lot to learn on how to use the genes for the goodness of mankind.

Keywords: Genetic Engineering, Gene Cloning, Biotechnology, Gene Therapy, Genetically Modified Organisms, CRISPR, CARS zygote

INTRODUCTION

Genetic engineering saw its infancy in early genetic practices orchestrated by renowned geneticist Gregor Mendel. Mendel's work began in the 18th century with the study of classical genetics and microbial genetics. These concepts begin to develop into more complex and advanced notions, one of which is genetic engineering. Specifically, genetic engineering's earliest experiments began in in the 1970s. At this time, genetic engineering encompassed the development of techniques and usage application. Genetic engineering is also known as *genetic manipulation, gene cloning, and recombinant DNA technology*. The term genetic engineering was coined in 1973 by American biochemists Stanley N. Cohen and Herbert W. Boyer. This term was originally reserved for the specified techniques used for modification of organisms through heredity and reproduction. The term ultimately developed a different means of application where the combining of DNA from multiple sources, within cells or invitro. These DNA molecules are then placed in host organism to generate. The basic foundations of genetic engineering are based on the principle that genetic information can be manipulated and used to provide a certain desirable outcome. In the late 1960s, genetic development and research had reached somewhat of plateau, primarily brought on by the lack of technological advancement in the field. At this time, all the knowledge needed to expand genetic research and manipulation was in place. Fortunately, by 1967, an enzyme called DNA ligase was able to be isolated. This process consisted of an enzyme joining two strands of DNA together. Subsequently, in 1970, the first restriction enzyme was isolated. Restriction enzymes work as molecular shears, sifting DNA strands at concise sequences. Following this development, the baseline requirements for the construction of recombinant DNA were present. Using the components of DNA ligase and restriction enzymes, scientists could then link the DNA of one organism to a completely different organism.

BENEFITS OF GENETIC ENGINEERING

Advantages of Genetic Engineering

Genetic engineering is a useful in biotechnology. Genetic engineering uses different techniques to alter genes of humans such as transformation and molecular cloning. Agriculture and medicine are two areas which make use of genetic engineering techniques most. The basic purpose of genetic engineering is to alter the genes. Those genes that are defective and do not work properly can be replaced by healthy genes using genetic engineering. Different enzymes are used to cut the DNA molecules from specific locations and are replicated by polymerase chain reactions and then utilized for various purposes. Some genes are not defective but increase the risk of certain disease or illness. E.G. The BRCA2 gene has been linked to breast cancer in many people of Spanish and Jewish decent. This gene may have positive benefits (not yet known), but most often is connected to an increased risk of Breast cancer. Medicine is looking into ways to remove this gene from the human genome, but for now we are just able to identify if a person has this gene and provide preventative procedures to mitigate the risk of breast cancer. Genetic engineering can have a positive change in human life. With these new developments in medicine and other areas, it is evident that genetic engineering has many benefits.

<u>The Human factor:</u> Genetic engineering opens new possibilities for biomedical enhancement requiring ethical, societal, and practical considerations to evaluate its implications for human biology, human evolution, and our natural environment. For instance, we can discuss the benefits of genetic engineering for human enhancement. There are multiple effects associated with germline heritable genetic intervention, we should consider the unit of impact to human populations and their natural environment and propose that a practicable distinction between 'therapy' and 'enhancement' is needed. There are several examples where technology has contributed to the lives of people by improving their inherent or acquired capabilities. For example, over time, there have been biomedical interventions attempting to restore functions that are deficient, such as vision, hearing, or mobility. If started from the time spectacles were developed, continuing in the last few years, with researchers implanting artificial retinas to give blind patients partial sight.

Recently, scientists have also successfully linked the brain of a paralyzed man to a computer chip, which helped restore partial movement of limbs previously non-responsive. In addition, synthetic blood substitutes have been created, which could be used in human patients in the future. The progress being made by technology in a restorative and therapeutic context could in theory be applied in other contexts to treat non-pathological conditions. Many of the technologies and pharmaceutical products developed in a medical context to treat patients are already being used by humans to 'enhance' some aspect of their bodies, for example drugs to boost brain power, nutritional supplements, brain stimulating technologies to control mood or growth hormones for children of short stature. Assistive technology for disabled people, reproductive medicine, and pharmacology, beside their therapeutic and restorative use, have a greater potential for human 'enhancement' than currently thought. There are also dual outcomes as some therapies can have effects that amount to an enhancement as for example, the artificial legs used by the South African sprinter Oscar Pistorius providing him with a competitive advantage.

The rapid advances in genetic technology makes this debate very topical. Moreover, genes are thought to play a very substantial role in biological evolution and development of the human species, thus making this a topic requiring due consideration. Currently, explorations are being conducted on how evolutionary biology could contribute to better assess the implications of human germline modifications. Overall, the aim is to contribute to the debate on human genetic enhancement by looking not only at the future, as it is so often done, but also at our evolutionary past.

Economic Benefits

The case study which we want to refer to here is how genetic engineering has resulted in a disease-resistant, high yield potato crop that mitigates the use of fungicides and saves humankind from many problems of the past. Some hardworking and dedicated scientists from the Netherlands have come up with this finding. Through various experiments and after multiple failed efforts, they had finally introduced a blight-resistant potato, which, if present during days of the Irish famine, could have turned tables around or might have stopped the problem before it even started. The “blight resistance gene” is inserted through genetic engineering into the genetic makeup of a regular potato to enhance its defense mechanism against fungal attacks (LYNAS). Before this, the solution to keep the blight at bay was "fungicide sprays," which increase the harvest and yield costs and have health side effects (LYNAS). Nevertheless, despite the numerous positive results of genetic engineering, just like the one mentioned above, "anti-GMO activists" always rise against such initiatives (LYNAS). Like in 2011, in Belgium, a trial field of blight-resistant potatoes planted by the scientist of Ghent University was destroyed by such activists. So, the genetic engineers and scientists have their share of problems and difficulties to go past while introducing such crops and advocating for them.

The genetically engineered blight-resistant potato, apart from being disease-resistant, has numerous other benefits as well. For instance, it could be "stored at colder temperatures for longer periods," which will vastly reduce the problem of surmounting food waste (Mulraney). “Pesticide residue” left inside potatoes after pesticide sprays will also not be a problem anymore (Mulraney). One great thing about genetic engineering is that it is not static but keeps on improving and growing. For instance, the earlier blight-resistant potato varieties due to modified composition did not taste that good. However, with enhanced modern research, good-tasting varieties have also surfaced. There is always some room for improvement. Besides, the diseases and pests, etc., also modify themselves and become resistant to prevalent anti-disease methods. Therefore, constant improvement is always the need of time. Flash forward to 2021, newer, more advanced blight-resistant potato varieties are being introduced. Farmers can earn more from them while doing away with the extra necessity of fungicide sprays. One research group investing in such potato varieties, headed by Mr. Gavin Powers, product manager of Agrico, estimates that most new varieties of blight-resistant potatoes can save up to “£400/ha” for growers (Farmers Guide). Also, as the population grows, conventional organic farming cannot fulfill the massive demand alone. Therefore, the future necessarily calls for blight-resistant genetically engineered potatoes to be introduced large-scale for their many benefits, which ought to reach humanity on a wide scale.

The happy news for the world is that the food administration and distribution authorities have started giving approvals to such excellent varieties of crops like blight-resistant potatoes. For instance, in the USA, one of the most technologically advanced countries in the world, the U.S. Environmental Protection Agency and the U.S. Food and Drug Administration have announced that genetically engineered potatoes are “safe for the environment and safe to eat,” and that these contain “only potato genes,” and no foreign injections (Ridler). This means that these potatoes are not transgenic, meaning thereby that they do not contain genes imported from some other plant or organism, etc. These potatoes' taste, texture, and nutritional qualities are also the same as conventionally grown potatoes, as per Doug Cole. He is the spokesman for Simplot, one of the leading companies working to yield genetically engineered potatoes.

Still, potatoes are one of the primary staple crops, standing at fourth place in consumption, after "corn, rice and wheat” (Ridler). Genetic engineering will not only reduce crop costs, and enhance storage time, and prevent disease, but newer varieties also prevent post-cooking health risks related to potatoes, about human health. For instance, genetically engineered potatoes have a significantly reduced amount of a naturally present chemical, which after being cooked at high temperatures, turns into a "potential carcinogen” (Ridler). However, even after such advanced progress, good work must continue to find even better varieties that can protect against more advanced varieties of blight. Hence, as far as the case study regarding the benefits of genetic engineering is concerned, in my eyes, this is one of the best examples of how human research and hard work can attain fabulous results. Furthermore, genetic engineering in the future and a solution to several problems presently facing humans and shall surface in the future. Therefore, rather than blinding criticizing scientists and researchers for working on such technology, one ought to have a logical debate about whether genetic engineering ultimately seeks to benefit humanity or cause it harm. Blight-resistant potatoes may be one good case study to start that healthy debate.

ISSUES WITH GENETIC ENGINEERING

Disadvantages of Genetic Engineering

Even though Genetic Engineering (GE) is a useful technology, it has many disadvantages that can alter humans and society forever. There are three main disadvantages that will be discussed thought out this section that will detail how genetic engineering can negatively affect businesses and customers. Firstly, Genetic Engineering can potentially introduce new diseases into society causing catastrophic consequences. Secondly, Genetic Engineering can cause crucial genes to become inoperative and dangerous genes that were once dormant to become active. Finally, Genetic Engineering can be copyrighted causing companies to potentially form a monopoly and have higher prices for other businesses as well as customers. Firstly, Genetic Engineering can potentially introduce new diseases into society causing catastrophic consequences. GE uses a process where different genes are isolated to target good genes or eliminate negative disease carrying genes. This process can have negative effects.

Table 1 shows the countries that are the leaders in genomics and the projects they have selcted to work on as apart of their research. Countries such as France and Japan are working on projects concerning rare diseases. There efforts are to isolate these rare diseas so that cures can be found or a way can be established to remove this gene from a persons DNA. The negative effects of this is that a rare disease carrying gene can be isolated and muitplies instead of being eliminated. This process can cause different pathogens to become more resistant than they normally would in a human. Charles Hagedorn , a specialist of Biotechnlogy, said “Genetically engineered organisims could potentially adversely impact both human health and the environment [as well as] pose risks that we simply do not know enough to identify” (Hagedorn 2000). Many of these events may be traced back to the cultivation and consumption of genetically engineered crops. Different dangers would be present with genetically modified animals, and they would be largely based on the new qualities put into the species, just as they are with plants. Transgenic crops could bring new allergens into meals that allergy sufferers are unaware of. One example is transferring the gene for one of the numerous allergenic proteins found in milk to plants like carrots. Mothers who know not to give their sensitive children milk are unlikely to know not to feed them milk-laced transgenic carrots. Genetic engineering is the only means for moving proteins across species boundaries into completely other animals. Scientists have limited ability to anticipate if a protein would cause food allergies in humans if taken. The only way to know for sure whether a protein will be an allergy is to try it. Importing proteins, particularly from nonfood sources, is thus a risky proposition in terms of allergenicity.

Table 1
Countries establishing themselves as leaders in genomics: select projects and their objectives

Country	Initiative	Objective
Australia	Australian Genomics Health Futures Mission	Develop national standards and protocols to enhance data gathering and analysis; promote the value of genomics to the broader community; and encourage government partnerships with philanthropists and businesses
China	100,000 Genome Project	Study how Chinese population transform from health to disease, environmental impacts, and the interactions between environmental factors and genes, and its influence on people's health
Estonia	Personalized Medicine Programme	Develop genotypes that will enable personalized reports for use in everyday medical practice through the national e-health portal
France	France Génomique 2025	Integrate genomic medicine into routine patient care and establish a genomic medicine industry to fuel economic growth. By 2020, France aims to have increased its annual sequencing capacity to 235,000 genomes, of which 175,000 are to come from cancer patients, and the remaining 60,000 from rare disease patients
Japan	Initiative on Rare and Undiagnosed Diseases	Develop innovative drug candidates by targeting novel, single pathological mutations, apply new NGS-based genome analyses to cases that remain unsolved, and facilitate international data sharing
Saudi Arabia	Saudi Human Genome Program	Study more than 5,000 inherited diseases using more than 10,000 samples from Saudi patients with inherited diseases that resulted in identification of more than 2,000 variants underlying the diseases
Turkey	Turkish Genome Project	Sequence the genomes of 100,000 Turkish nationals and increase that number to 1 million genomes by 2023
United Arab Emirates	United Arab Emirates—Dubai Genomics	Sequence all of its 3 million residents. Dubai Genomics is one of numerous projects within the Dubai Future Foundation's "Dubai 10X Initiative," launched to catapult the UAE 10 years ahead of the rest of the world
United Kingdom	100,000 Genome Project	Incorporate genome sequencing in routine healthcare through the Genomic Medicine Service (GMS). Sequenced 71,095 whole genomes
United States	All of Us Research Program	Glean health and wellness data from 1 million or more Americans

In genetic engineering, antibiotic resistance genes are widely utilized as "selectable markers." These markers are used to detect cells that have taken in foreign genes early in the engineering process. Even after the genes are no longer needed, they are still expressed in plant tissues. In genetically modified plant foods, many antibiotic-resistance genes are fully functioning.
Food containing antibiotic resistance genes may have two unfavorable outcomes. First, consuming certain foods while taking antibiotics may reduce their effectiveness in fighting disease. Antibiotic resistance genes produce enzymes that can break down antibiotics. If you eat an antibiotic with a raw tomato that carries an antibiotic-resistance gene, the medication may be destroyed in your stomach. Second, antibiotic resistance genes could be passed along to human or animal illnesses, making them resistant to antibiotics. If this happens, the already grave problem of antibiotic-resistant disease organisms will be exacerbated. Even though unmediated transfers of genetic material from plants to bacteria are extremely rare, considering the importance of antibiotic resistance, any possibility that they may occur must be thoroughly investigated (Scully, 2003).

A diverse spectrum of creatures can produce toxic compounds. These chemicals are used by plants to defend stationary species from the numerous predators that exist in their environment. In some cases, plants can contain inactive pathways that lead to dangerous chemicals. The introduction of additional genetic material through genetic engineering could reactivate these dormant pathways or increase the amount of harmful compounds in the plants in other ways.

This may happen, for example, if the on/off signals of the inserted gene were placed on the genome in areas where they could activate previously inactive genes. Heavy metals such as mercury can be removed from the soil and concentrated in plant tissue thanks to some of the new genes being added to crops. The goal of growing these crops is to make it possible to use municipal sludge as a fertilizer. Sludge provides valuable plant nutrients, but it is frequently polluted with poisonous heavy metals, preventing it from being utilized as fertilizer. The objective is to modify plants so that those metals are removed and sequestered in non-edible portions of the plant. Metals are sequestered in the roots of tomatoes and the leaves of potatoes, for example. Turning on genes in only select portions of the plants necessitates the employment of genetic on/off switches that are activated only in specific tissues, such as the leaves.

If the on/off switches in edible tissues are not completely turned off, excessive levels of dangerous metals may contaminate foods. There are additional environmental risks associated with the processing and disposal of metal-contaminated plant fragments after harvesting. Although the bulk of health risks are produced by the introduction of additional genetic material into organisms, the removal of genes and gene products can also be problematic. For example, genetic engineering might be used to generate decaffeinated coffee beans by deleting or turning off genes involved in caffeine production. Caffeine, on the other hand, helps to keep fungus away from coffee beans. Beans that are unable to produce caffeine may become infected with fungus that produces toxins. Aflatoxin and other fungal toxins are very toxic to humans and can survive food preparation. As with any new technology, the whole range of risks associated with genetic engineering is probably certainly unknown. Our ability to imagine what can go wrong with technology is limited by our current lack of understanding of physiology, genetics, and nutrition.

Consider the likelihood that genetically modified plants will become weeds, as well as the potential environmental impact. Any plant that develops in an area where humans do not want it is referred to as a weed. The word spans anything from Johnson grass strangling crops in fields to kudzu blanketing trees to melaleuca trees invading the Everglades. The plants are growing unsupported in each case in regions where they are generating problems. In agriculture, weeds can dramatically diminish crop yield. In unmanaged areas like the Everglades, invasive trees can displace local vegetation and destroy entire ecosystems. It is not assured that novel genes inserted into crops will remain in agricultural fields. If relatives of the altered crops grow adjacent to the field, pollen from those plants can swiftly transfer the new gene to those plants. The new characteristics could allow wild or weedy relatives of crop plants to persist in harsh conditions, effectively rendering them weeds in the classic sense. A gene that alters the oil composition of a crop, for example, could spread to nearby weedy relatives, allowing the seeds to survive the winter with the altered oil composition. Overwintering can turn the plant into a weed or exacerbate whatever weedy traits it already has (Fasani et al., 2018).

Genetically engineered crops that are resistant to chemical herbicides are intrinsically linked to the use of certain chemical pesticides. The chemical herbicide mix used across the country may change because of the adoption of these crops. These altering patterns may result in increasing environmental effect to the extent that chemical herbicides differ in their environmental toxicity. Additionally, widespread use of herbicide-tolerant crops may speed the evolution of herbicide resistance in weeds, either as a result of higher chemical exposure or as a result of the herbicide trait being passed down to weedy relatives of crops. Introducing foreign genes to plants could have disastrous consequences for mammals in a variety of settings. Mice and deer could be harmed if they eat crops or crop debris left in the fields after harvesting if crop plants like tobacco or rice are genetically modified to produce polymers or pharmaceuticals. If fed to other fish, fish-eating birds, and mammals, fish modified to include metal-sequestering proteins (such fish have been proposed as living pollution clean-up devices) could be detrimental to other fish, fish-eating birds, and mammals. One of the most common uses of genetic engineering is the development of virus-tolerant crops. Viral components are inserted into plant genomes to develop these crops. For unexplained reasons, plants that manufacture viral components on their own are immune to later infection by those viruses. Recombination and transpeptidation, on the other hand, can result in the development of new or worse viruses in such plants (Hollander, 2018).

Current Limitations and Shortcoming

"Nothing vast enters the life of mortals without a curse." — Sophocles
Humankind has always evolved new methods to practice medicine that have either brought a great number of benefits or terrible consequences that often pose ethical issues. With the rise of the technological revolution there has been a new method in the medical field in which scientists are now able to edit the human genome. In the 1970s, scientists developed several new methods of changing "hereditary material of living organisms." With this new technological development, genes can be introduced to cells, or even further, into whole organisms--permanently altering the organism. Likewise, genes of an organism can be altered and replaced. This technological practice is called genetic engineering, and it has the potential to do great harm to the human population, but some would argue that it can bring vast benefits to the species. As a result, it is a great ethical concern as to whether scientists should pursue the use of this technology.

One of the main limitations and opportunities for genetic engineering is that it can be easily abused. An example of this is the newest gene editing technology CRISPR, which stands for: Clustered, Regularly Interspaced, Short Palindromic Repeats. This technology is very easy to use and a lot cheaper than previous technologies on gene editing. CRISPR allows scientists to manipulate the sequence of DNA in a very precise manner with less of the number of resources needed of previous technologies on gene modification. According to a thought-provoking TED talk by Biologist Paul Knoepfler, he estimates that within fifteen years scientists could use the gene editing technology CRISPR to make certain "upgrades" to human embryos, also known as the popular term "designer babies." With the knowledge and fear in mind, he ultimately argues that we should not allow for the genetic modification of people, because "it is too dangerous and too unpredictable." This point is further emphasized by specialists on medical ethics at New York University's Langone Medical Center. Dr. Arthur Caplan during an interview conducted by Vice News reporter Isobel Young exploring the moral and ethical concerns behind the fast-developing gene editing technology. "The biggest ethical concern of all is that we are going to try it before we understand it" says Dr. Caplan. Which is why the evolution of CRISPR technology led many scientists to come together from several developed countries to pass a moratorium until further research is proven that there are less risks and further damage after gene editing.

With the emergence of CRISPR's technology efficiency of gene editing, there is the opportunity to modify human traits and even go as far as producing babies that are genetically enhanced. In theory, with CRISPR's technology one can alter and determine aesthetic attributes, intellects, physical ability and so on. Researchers and ethicists from the International Summit on Human Gene Editing have written and agreed that "until germline genome editing is deemed safe through research, it should not be used for clinical reproductive purposes; the risk cannot be justified by the potential benefit" (Lanphier, Urnov, Haecker, Werner, & Smolenski 2015). Other researchers have also argued that "there may never be a time when genome editing in embryos will offer a benefit greater than that of existing technologies, such as preimplantation genetic diagnosis (PGD) and in-vitro fertilization (IVF)" (Lanphier et Al. 2015). As always in the field of science and medicine, there will be someone who will push the boundaries and experience with the technology. In fact, despite "all ethical debates and laws, a group of Chinese scientists have published a paper in April 2015 on the use of CRISPR-Cas9 on human non-viable embryos. Their main objective was to find a cure for beta-thalassemia, a hereditary blood disorder, but their experiment could not develop into human fetus. A year later another Chinese team used CARS zygote to induce mutation on genes to make zygotes immune to HIV Virus" (Omodamilola, 2018). Although CRISPR technology "is very simple, easy to use and cheap unlike the previous gene editing techniques such as Transcription activators-like effective nucleases (TALENS), it can also be used to knock out a gene and replace it with another gene for disease therapy" (Zhang 2015). In 2016 the same techniques that were prohibited from using the CRISPR on human genes or designing babies were used to engineer the human immune system to fight against cancer.

According to an article from the Journal of Biomedical and Pharmaceutical Sciences titled *CRISPR Technology Advantages, Limitations and Future Direction* author: Omoyayi Ibrahim Omodamilola from the Department of Biomedical Engineering explained that "the effect of off-target can alter the function of a gene and may result in genomic instability, hindering its prospective and application in clinical procedure. A Single guide RNA also known as chimeric RNA is the combination of CRISPR RNA (CrRNA) and trans-activating RNA (TrRNA). SgRNA's 20 base nucleotides are complementary to the target DNA of interest with a trinucleotide known as Protospacer adjacent motif (PAM) adjacent to sequence which is mostly NGG (where N can be Adenine, Guanine, Cytosine or Thymine)" (Omodamilola, 2018). Finally, Economic Disparity is also a big shortcoming and potential limitation to mass genetic engineering access. Some possible unanswered questions will eventually be answered are: How should we prevent the use of CRISPR technology of gene editing from only being accessible to the wealthy and increasing existing disparities in access to health care and other interventions? According to Robert Truog, director of the Center for Bioethics at Harvard Medical School "Aside from the safety risks, human genome editing poses some hefty ethical questions. For families who have watched their children suffer from devastating genetic diseases, the technology offers the hope of editing cruel mutations out of the gene pool. For those living in poverty, it is yet another way for the privileged to vault ahead. One open question is where to draw the line between disease treatment and enhancement, and how to enforce it, considering differing attitudes toward conditions such as deafness" (Todd Bergman, 2019). The genetic engineering of humans poses major ethical dilemmas and shortcomings due to lack or research and strong boundaries. However, the benefits of this technology could be to eradicate certain health issues that humanity have yet to find cures for. Conversely, this technology could be the start of a new eugenics' movement.

FUTURE IMPLICATIONS/IMPROVEMENTS OF THE TECHNOLOGY

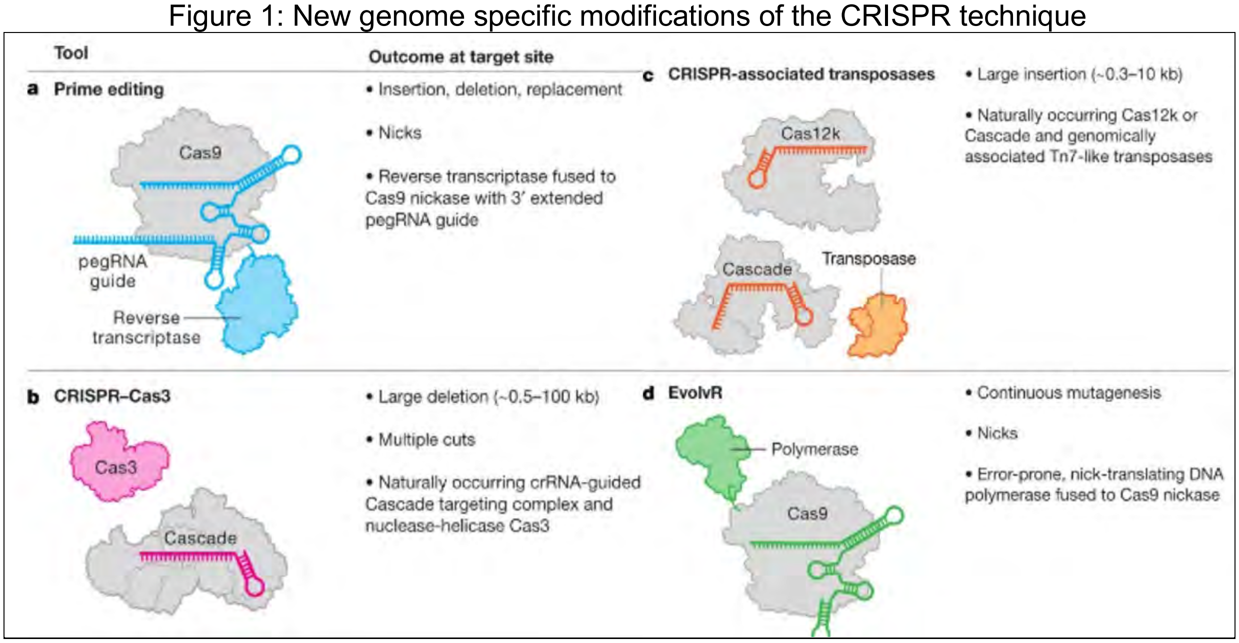
Academia continues to build upon prior knowledge that has advanced the science of genetic engineering. Future perspectives on the subject continue to tackle the obstacles in determining how this tech can be of benefit to society. Throughout the years the tools and techniques have been refined by scientist, collaborating their processes within the science community. These collaborations secure that future tools and techniques will have improved accuracy in the field of genetic engineering. Emerging technology such as CRISPR has set the stage for growth in gene editing but also shed light on many unanswered questions that the science poses, such as the health factor risk to that comes with human trials and the minimization of patient's exposure to genotoxicity during gene editing techniques. Public perception that challenges the ethics of the science is another obstacle that will have to be overcome but the extensive trial periods that are continuously being conducted and the growing relevance of gene editing technologies may sooth the public's opinion on the field. Trends that lead the way in genetic engineering still have some hurdles to overcome before reaching the level of technology out of science fiction but the progress being made does show promise in making effective change in the world's medical field, which translates to bettering the lives of people across the globe.

TECHNIQUES/TECHNOLOGIES

Gene Editing

A recent advancement in genetic engineering comes in the form of a method called gene editing. "Gene editing refers to a group of techniques where specific genetic modifications are targeted in the native genome of a given organism (Osmand & Colombo, 2019)." A famous advancement of the gene editing technology has come in the form of the CRISPR/CAS technique. CRISPR shows promise in potentially benefiting not only the medical field but also: provide fuel alternatives, improve crop yield, and contribute to food security (Osmand 2019). The adaptability into different fields is attributed to swift development of tools and leniency being established to conduct trials to test these technologies. Future techniques could allow scientist to home in on specific gene behaviors like Columbia assistant professor, Alex Chaves, MD, PHD, who's research involves the study of Alzheimer's disease and activating and silencing specific genes in the disease (Tregaskis, 2020). New modifications of the CRISPR technique shown in *figure 1* highlight the genome specific modifications: "a, Cas9 binds to and nicks the genomic target, after which the reverse transcriptase copies the sequence of the prime-editing guide RNA (pegRNA) to the target site. b, Cascade binds to a genomic target, inducing processive cleavage by Cas3 and generating large deletions. c, Cascade or Cas12k binds to the genomic target and directs donor DNA insertion by the Tn7-like transposase. d, Cas9 binds to and nicks the genomic target, after which the error-prone polymerase generates diversity in an adjacent window, thus enabling directed evolution (Dounda, 2020)."

Figure 1: New genome specific modifications of the CRISPR technique



With rapid innovation of gene editing technologies many science labs are beginning to run more complex trials requiring more complex organisms than the usual lab mice. The question of patient safety, as more human trials are being conducted, is at the forefront of potential challenges along the way. Proper understanding of the essential biology of this technology will be needed to protect patients. Modifying such small structures in biology could lead to undesired results if editing is applied to an incorrect region. An example is seen in CRISPR technology, which, at times, has yielded off-target effects in testing (Hirakawa, 2020). For future application, these errors must be minimized to ensure delivery to targeted genetic sites.

Gene Therapy

An alternative genetic engineering technique that has also gained some attention in the medical community is the technology of gene therapy. The practice involves "the administration of foreign genomic material into the host tissue to modify the expression of a gene product or to change the biological properties of cells for therapeutic use (Belete 2021)." Current use of this technology has aided scientist in the treatment of various forms of cancers and other diseases in patients. Building on current applications of gene therapy in Rheumatoid Arthritis treatment shows potential direction for the technology. Observing the relatively poor regeneration of mammalian cartilage in between joints while other vertebrates can completely regenerate limbs; singling out specific genomes of these vertebrates could lead to specific expression of similar genomes in humans to aid in Rheumatoid Arthritis treatment (Deviatkin, 2020). A risk highlighted in trails has been the genotoxicity threat of cells that have had genome therapeutic procedures. Genotoxicity refers to when a substance has the capability to damage gene structures of cells. The modification process of cells runs the risk of exposing patients to harming effects if not completed correctly. The safety precautions to allow for human clinical trials so most studies to combat the effects of genotoxicity have been conducted on mice (Ohmori,2020). The current block in furthering gene therapy technologies is in how scientist will learn to apply trial runs that protect human subjects.

CONCLUSION

The horizon for genetic engineering has the potential to branch out into various areas in future application. The science communities' collaboration with these technologies and techniques has accelerated the output of practical functionality in medical procedures and identify the areas of the field that need improvement. Through continued studies genetic engineering innovation has potential to secure a promising future in medicine.

REFERENCES

References available upon request.

DECISION SCIENCES INSTITUTE

Youth Narratives in Relation to Trauma

Audra Dankwardt and Jesslynn Rocha Neves
UMass Amherst
adankwardt@umass.edu jrochaneves@umass.edu

ABSTRACT

This paper attempts to understand and recognize the powerful nature of narrative storytelling. In particular, it seeks to utilize the healing aspects of storytelling for youth who have experienced trauma. By looking at the prevalence of adverse childhood experiences across groups, it recognizes the disproportionate levels of trauma for minoritized youth, including youth of color and low-income families. Storytelling is ingrained in humanity, and the need for community and agency spans across all experiences and identities. Narrative storytelling, with structure and support, can help youth connect with, create meaning from, and heal from their negative experiences.

KEYWORDS: Trauma, Storytelling, Healing, ACEs, and Youth

INTRODUCTION

Everyone has a story: a collection of experiences and important moments that make them who they are. These stories start in our childhood and are often shaped by the people around us. A healthy child, without adverse childhood experiences, might be able to tell a story of a happy and loving family, peer friendships, and big accomplishments. These narratives are important because they allow us to create a sense of purpose and unity across time and experiences (Adler 2012). Unity, as a concept, is incredibly important to our identity because it allows us to integrate our past, present, and future selves, ensuring we feel connected to ourselves and our experiences (Adler 2012). More than anything, it allows us to create meaning from what has happened to us and allows us agency in our decisions (Adler 2012). Narrative storytelling is the process where individuals can share the culminations of these moments in their lives that created their identity – either to themselves or others.

LITERATURE REVIEW

Literature Review of Topic 1

Narrative storytelling is not just an individual process, and the world around us constructs stories of our experiences, cherry-picking which details to include. Master narratives are stories that are central to a culture and attempt to explain why one group is superior to another (Wing, 2021). They are often also supported by the state and government in the form of history books, education, and the legal system (Wing, 2021). Counter narratives, on the other hand, attempt to create an alternate reality that shatters complacency and challenges the status quo (Wing, 2021). These narratives are important because they often come from historically silenced groups and provide the other side of the story. A common example of a dominant and counter narrative is that of Christopher Columbus and the Arawak people. In the American education system, we are taught the story of colonization through the stories of Columbus and his men – often completely disregarding the Arawak people's stories entirely. It is incredibly important to empower minority groups to share their stories that are silenced and ignored – both so they can construct their own identities and narratives, and also so that people can recognize the reality of the situation (a much more complicated history).

Literature Review of Topic 2

In order to determine who experiences trauma and at what rates, I examined the prevalence of ACEs among children. Adverse Childhood Experiences (ACEs) are negative events that occur during key stages of development (Rocha Neves 2023). These events can be potentially traumatizing to youth and can range from parental divorce to sexual abuse. Without intervention, these ACEs can create toxic stress for youth who are not equipped with resources to handle them (Rocha Neves 2023). Moreover, ACEs are very common with almost half (45%) of all youth experiencing at least one (Sacks and Murphey, 2018). For many children (1 in 10), they will experience three or more ACEs; this places them at "especially high risk" for negative outcomes (Sacks and Murphey, 2018).

Unfortunately, ACEs happen at much higher rates for youth of color compared to white youth. For hispanic children, 51% have experienced at least one ACE – and for black non-hispanic children, this number rises to 61% (Sack and Murphey, 2018). On the other hand, 40% of non-hispanic white children and only 23% of Asian non-hispanic children experience ACES (Sacks and Murphey, 2018). These ACEs can contribute significantly to negative health outcomes including alcoholism, drug abuse, depression, suicide, and obesity (Sacks and Murphey, 2018). Additionally, ACEs are associated with lower educational attainment, unemployment, and poverty (Metzler et al. 2017).

Table 1
Prevalence of Individual ACEs for Children in various Racial/Ethnic groups

	Hard to cover basics like food or housing somewhat or very often	Parent or guardian divorced or separated	Parent or guardian died	Parent or guardian served time in jail	Saw or heard parents or other adults slap, hit, kick, or punch in home	Victim of or witness to violence in neighborhood	Lived with anyone mentally ill, suicidal, or severely depressed	Lived with anyone with problem with alcohol or drugs
United States								
White, NH	22	23	3	7	5	3	9	10
Black, NH	37	35	7	16	9	7	6	8
Asian, NH	14	7	2	1	2	2	2	1
Other race, NH	31	27	4	11	7	6	11	12
Hispanic	29	28	3	8	6	4	6	9
All children	**25**	**25**	**3**	**8**	**6**	**4**	**8**	**9**

* Note: Yellow shading = Percentage is higher than white non-Hispanic children at a statistically significant level. Blue shading = Percentage is lower than white non-Hispanic children at a statistically significant level. Red shading = Estimate should be interpreted with caution, because the relative confidence interval is greater than 120 percent. (Sacks and Murphey, 2018).

Literature Review of Topic 3

Adverse Childhood experiences also make youth more susceptible to becoming justice involved (youth involved with the criminal justice system). For justice involved (JI) youth, the rates of ACEs are much higher, which can affect their trajectory in the criminal justice system (Bergquist et al., 2022). These findings are also compounded by the fact that youth of color are far more likely to have contact with the criminal justice system (Kempf-Leonard, 2007). However, studies have found that protective factors, such as future orientation, can mitigate the effects of ACEs and help prevent youth from delinquent activity (Chen et al., 2013). Children who experienced multiple ACEs require positive adult relationships, social support, and school-connectedness to prevent delinquency for at-risk youth (Brown and Shillington, 2017), (Davis et al., 2019). I also want to recognize how traumatizing the criminal justice system can be, especially for people of color. For these minoritized youths, the cycle (trauma → justice involved → trauma from the criminal justice system) can be difficult to break.

YOUTH NARRATIVES IN RELATION TO TRAUMA

For children who experience trauma, their past experiences can muddle their identities – making it difficult for them to recognize who they are outside of their abusers, environment, or pain. While trying to make sense of their experiences, attempting to convey them to outside parties can become impossible. Children of minority groups also experience trauma at much higher rates, making it increasingly difficult to share their counter narratives to the master narrative. Standing up to the American government and education system can feel like an insurmountable task, as they struggle to share how they were wronged. Youth Narrative Storytelling, therefore, is essential for children to heal from their trauma.

For youth struggling to separate themselves from their trauma and develop their identity, it can be helpful to find collective memories – those shared by other people with similar experiences. Friendships forged from this storytelling create attachments that are vital to survival for youth who experience trauma (Mitchell and Selfridge, 2018). Moreover, these stories continue to move and transform as youth construct spaces for storytellers and their audience to find collaboration in the narrative process (Mitchell and Selfridge, 2018). Storytellers also use stories as a means for "testing the waters" and recognizing how safe people or places may be (Mitchell and Selfridge). These spaces and communities built from storytelling, allow youth to recognize who they were and who they are now – seeing themselves in the stories others share (Mitchell and Selfridge). The trust, safety, community, and friendship built from these stories encourage youth to face their trauma and acknowledge how it has impacted their lives – letting them begin the healing journey.

Often, stories can be misunderstood, as humans do not have a universal communication model for trauma narratives. Children struggling trying to find the words to describe how they are feeling internally may use behavior to express the overwhelming and enduring memories of traumatic experiences, yet adults regularly interpret this behavior as "acting out" and seek to remedy or "fix" the behavior (Regan, 2020). Instead of trying to eliminate this behavior, parents should work to understand and hear the narratives. This is because youth need to feel like someone else is taking on their pain in order to find the capacity to recover from their trauma (Regan, 2020). Constructing a space where narrative storytelling is acceptable (and providing assistance with language) can help youth find the understanding they desire from adults and educators in their lives.

Since youth need to tell their stories, but often struggle with how to communicate their complex emotions and feelings, it is beneficial for adults in their lives to provide recommendations and structures. Helping children find the language that best matches their feelings is one of the first steps in teaching them about narrative storytelling. While "finding the words," people should first assess the memories, verbal abilities, and attention spans of the youth they wish to help (Key and Newland, 2020). Next, they can begin identifying the format (verbal, written, visual, etc) and events of focus, especially if a child has experienced multiple traumas (Key and Newland, 2020). Clinicians should be available during this process to help youth stay focused on the moment, organize their thoughts and memories, and maintain emotional regulation (Key and Newland, 2020). While these narratives are not always formal, structure surrounding the creation of narratives and repetition can help youth to regulate themselves and maintain connection to their past/trauma.

Strengths of Narrative Storytelling

Youth often feel as though their choices and decisions make no difference for their fate. They can feel disempowered by the adults in their life who seek to control the outcome of their life. Decisions are limited, preferences are ignored, and opinions are rarely asked. Narratives are important to youth because they allow youth minorities to sustain their sense of agency among their group, in a world that seeks to disempower them (Jackson 2002). By analyzing their experiences, youth can attempt to understand and share how to change outcomes. Collectively sharing positive stories and advice allows youth to connect and teach each other helpful strategies to combat oppression (Lopez 2020). Empowering youth helps them to change their circumstances and limits learned helplessness.

For many youth, they find strength in sharing collective experiences and memories: creating empathy and unity when they feel otherwise alone. Storytelling allows youth to reflect on their position, their identity, and how their decisions impact the way they are viewed by people around them (Mitchell and Selfridge 2017). These stories take place everyday and change based on the audience, environment, and context (Anderson 2010). Furthermore, youth use these stories to form relationships with others, relying on active engagement between listener and teller (Sabiston 2013). Often youth can feel alone with their trauma, accepting blame for the experience and isolation as the consequence. Youth use these stories to find commonality between themselves and peers, especially in circumstances of environmental trauma. For someone who endures a community trauma – such as a pandemic, environmental disaster, or mass shooting – finding community can come from neighbors, peers, and family members. At this time, they also find community on social media, sharing their stories for others to relate to. For some youth, especially street youth, it can be challenging to find people who have similar experiences of trauma. Therefore it is essential for youth to find effective and meaningful ways to share their stories, and to heal through collective trauma and community.

Weaknesses of Narrative Storytelling

Coming to terms with trauma and expressing this pain with another person is a courageous act. People share their narratives because they need to feel understood and listened to – they need to feel loved. However, to be met with silence instead of comfort, or disagreement instead of validation, can be disheartening and retraumatizing. Historic and cultural trauma can be especially challenging to share, as licensed professionals pathologize the response of the individual and attempt to "diagnose" the problem (Jackson, 2003). Instead, they should act as the audience, listening to the stories and only offering management techniques. Narrative storytelling needs to be about trustworthiness and transparency, collaboration and mutuality, empowerment and choice/control, and safety (Sweeney et al., 2018). Mutuality and power-sharing are especially important to combat the helplessness and shame people may experience if they feel dependent and controlled when they are being “helped” (Blanch et al., 2012). Victims need to be able to control their own road to recovery and tell their narratives. “Helpers” should be partners that offer resources or techniques, but never dictate actions or feelings.

While often informal, narrative storytelling can come with the expectation of consequences for the offender. For example, some victims must recount their stories in criminal proceedings or in lawsuits in order for courts to determine a verdict. These accounts can be especially difficult because they often require the teller to dissociate from the event and share the “facts” of a case with certainty (Burchard, 2019). These forms of narrative storytelling are not healing and do not provide accommodations or recognize the additional burden placed on the storyteller (Burchard, 2019). Testimonials from victims can come with the expectation of justice (often not accomplished), but do not promote understanding and recovery (Burchard, 2019). Many youth may recount their experiences to people in power (ex. educators) in the hopes that they can stop the pain. However, educators and mandated reporters often are not adequately trained and feel as though child protective services do more harm than good to the child (Figueroa and Maldonado, 2020). Many educators also believe child protective services will not investigate the claims (Figeuroa and Maldonado, 2020). With educators identifying a significant lack of faith in the criminal justice system, we cannot expect children (especially minoritized children) to trust that same system.

IMPLICATIONS FOR PRACTICE AND RESEARCH

Youth Narrative Interventions to Support Healing

Clinicians often require investigative interviews from clients to determine a prognosis and plan for recovery. With unresolved trauma, memories and descriptions of traumatic events can be incoherent, fragmented, or contradictory (Bifulco, 2021). Clinicians can assist victims in this process by scaffolding questions, however, if memory encoding was impacted by trauma, memories can be difficult or impossible to retrieve (Bifulco, 2021). For this reason, it is important that people working with youth recognize that narratives do not need to be complete and may require assistance to construct. This process should be a partnership between the teller and listener, and must include careful consideration to the impact trauma has on memory. Additionally, awareness of the ways that historical and cultural oppression may impact long-term memories is crucial. Therefore, interventions without cultural competence and recognition of systemic oppression will not be productive and could cause further trauma.

Storytelling can look different for every individual, but there are a few identifiable components that promote survival and healing. In order for radical healing to happen across groups, the story must include: storying influences, mechanisms of storying survival, content of storying survival, context of storying survival, and impact of storying survival (McNeil-Young, 2023). In this case, storying survival refers to using narratives in the promotion of liberation from racial trauma (McNeil-Young at al., 2023). Interventions, therefore, can include assistance in identifying these components, especially for minoritized youth. For individuals struggling with content due to memory loss, they can focus on the influences, mechanisms, context, and impact. These components are interconnected, so these parts should not be referred to in isolation. Rather, treatment includes partnership between components where parts may compensate for failure in others.

Future Research

Given more time and resources, including the ability to interact with participants, future research could evaluate the effectiveness of storytelling practices in school. These practices could be informal, such as a student making off-hand comments to a teacher, or formal, such as a counselor talking to a student about their history. Given that many mandated reporters identified insufficient training as a barrier to providing effective support, I think narrative storytelling efforts within schools could be beneficial to healing beyond the criminal justice system. Furthermore, since a disproportionate number of traumatized youth are minorities, narrative storytelling could be a helpful intervention to students who lack resources for formal care.

CONCLUSION

Informal narrative storytelling has existed for generations as a way for humans to feel understood and cared for. Organizations and communities have twisted these narratives to promote status quo – deeming one group inferior through exclusion or denial of experiences. Therefore, alternate narratives were created as a means of survival for oppressed and silenced groups. These narratives can help individuals to create collective memories that provide agency and limit feelings of loneliness. These narratives can be involuntary – in the form of behavior – which can cause retraumatization if listeners ignore signals and dismiss claims. To provide effective interventions and promote positive experiences in narrative storytelling, helpers should recognize the impacts of historical oppression and trauma on memory consolidation. Finally, storytelling components are interconnected, but should include influence, mechanisms, content, context, and impact in order to promote liberation for oppressed groups. With this in mind, storytelling as a practice in schools could be beneficial only if educators engage in the learning process that goes into healing through narrative storytelling.

REFERENCES

Adler, J. M. (2012). Living into the story: Agency and coherence in a longitudinal study of narrative identity development and mental health over the course of psychotherapy. Journal of Personality and Social Psychology, 102(2), 367–389. doi:10.1037/a0025289

Anderson, K. E. (2010). Storytelling. 21st Century Anthropology: A Reference Handbook, 277–286. https://doi.org/10.4135/9781412979283.n28

Біфулко, А. (2021). Childhood trauma in women and fragmented interview narratives – some interdisciplinary methodological and clinical implications. East European Journal of Psycholinguistics, 8(1). https://doi.org/10.29038/eejpl.2021.8.1.bif

Blanch A, Filson B, Penney D, et al. (2012) Engaging Women in Trauma-Informed Peer Support: A Guidebook. National Centre for Trauma-Informed Care.

Brown, S. M., & Shillington, A. M. (2017). Childhood adversity and the risk of substance use and delinquency: The role of protective adult relationships. Child Abuse & Neglect, 63, 211– 221. https://doi.org/10.1016/j.chiabu.2016.11.006

Burchard, M. (2019). Non-standard witness: Lessons from working with/for traumatized children. Testimony and Trauma, 128–152. https://doi.org/10.1163/9789004391130_008

Chen, P., Voisin, D. R., & Jacobson, K. C. (2013). Community violence exposure and adolescent delinquency: Examining a spectrum of promotive factors. Youth & Society, 48, 33–57. https://doi.org/10.1177/0044/118X12475827

Davis, J. P., Ports, K. A., Basile, K. C., Espelage, D. L., & David-Ferdon, C. F. (2019). Understanding the buffering effects of protective factors on the relationship between adverse childhood experiences and teen dating violence perpetration. Journal of Youth and Adolescence, 48, 2459–2359. https://doi.org/10.1007/ s10964-019-01082-3 Figueroa, M., & Maldonado, M. (2020). Educator's Barriers to Reporting Child Maltreatment. CSUSB ScholarWorks.

Jackson, Michael. The Politics of Storytelling: Violence, Transgression, and Intersubjectivity. Museum Tusculanum P, 2002.

Jackson V (2003) In our own voice: African-American stories of oppression, survival and recovery in mental health systems. Off Our Backs, 33(7/8): 19–21.

Kempf-Leonard, K. (2007). Minority youths and juvenile justice. Youth Violence and Juvenile Justice, 5(1), 71–87. https://doi.org/10.1177/1541204006295159

Key, D., & Newland, R. (2020). Creating trauma narratives for young children. The Brown University Child and Adolescent Behavior Letter, 36(10), 1–6. https://doi.org/10.1002/cbl.30493

Lopez, R. M. (2020, March 23). Discursive De/Humanizing: A Multimodal Critical Discourse Analysis of Television News Representations of Undocumented Youth.

McNeil-Young, V. A., Mosley, D. V., Bellamy, P., Lewis, A., & Hernandez, C. (2023). Storying Survival: An approach to radical healing for the black community. Journal of Counseling Psychology, 70(3), 276–292. https://doi.org/10.1037/cou0000635

Metzler, M., Merrick, M. T., Klevens, J., Ports, K. A., & Ford, D. C. (2017). Adverse childhood experiences and life opportunities: shifting the narrative. Children and Youth Services Review, 72, 141-149. Retrieved from http://www.sciencedirect.com/science/article/pii/S0190740916303449

Mitchell, Lisa, and Marion Selfridge. Because That's What We Do . . . We Sit and We Drink and We Talk': Stories and Storytelling among Street-Involved Youth. Jeunesse: Young People, Texts, Cultures, vol. 9, no. 2, 2017, pp. 91–111., https://doi.org/10.1353/jeu.2017.0026.

Regan, Dominic. How the Challenging Behavior of a Traumatized Child Tells Their Trauma Story and Is a Vital Part of Their Recovery. Journal of Social Work Practice, vol. 34, no. 1, 2019, pp. 113–118., https://doi.org/10.1080/02650533.2019.1572081.

Sabiston, Les. Native Youth and the City: Storytelling and the Space(s) of Indigenous Identity in Winnipeg. M.A. thesis, U of Manitoba, November 13, 2013, mspace.lib.umanitoba.ca/bitstream/handle/1993/22257/Sabiston_Les.pdf?sequence=1. Sacks, V., & Murphey, D. (2018). The prevalence of adverse childhood experiences, nationally, by state, and by race or ethnicity. Bethesda, MD: Child Trends. Stephens-Davidowitz, S

Sweeney, A., Filson, B., Kennedy, A., Collinson, L., & Gillard, S. (2018). A paradigm shift: Relationships in trauma-informed mental health services. BJPsych Advances, 24(5), 319–333. https://doi.org/10.1192/bja.2018.29

Wing, L. (2021) Lecture Three: The Master Narrative [Discussion] retrieved from Umass Amherst Legal 482

DECISION SCIENCES INSTITUTE

Zooming in on Inflation and Associations with Market Returns

Raymond Shao
William Mason High School, Mason, OH
Email: rayyu.shao@gmail.com

ABSTRACT

We investigate different measures and components of inflation and their roles in predicting market excess returns. We dissect consumer price index (CPI) and personal consumption expenditures price index (PCEPI) by energy, food, and core, and examine corresponding inflation rates. We apply neural network/deep learning and investigate the predictive performance. We find that inflation rate based on the CPI's food component, along with stock variance and scaled market price, has 5% monthly out-of-sample R^2 in predicting S&P 500 excess returns. For a single-factor model, the inflation rate on CPI's energy component alone has more predictive power than the core inflation rates.

KEYWORDS: CPI, Deep Learning, Inflation Rate, Machine Learning, Stock Return

INTRODUCTION

Inflation plays a profound role in predicting market returns. Recent years' high inflation rates have greatly impacted the stock market, the Federal Reserve's decision to increase the Fed fund rate, and consumers in general. In economics, inflation refers to an increase in the general price level of goods and services in an economy. And inflation rate (INF) is often measured by the change in the Consumer Price Index (CPI), which tracks the prices of a basket of goods and services commonly purchased by urban consumers. The inflation rate based on CPI measures the rate at which the consumers' cost of living is increasing or in rare cases, decreasing. The Personal Consumption Expenditures Price Index (PCEPI) is an alternative inflation measure. In addition to the prices paid by consumers for CPI, the PCEPI adds prices paid by businesses and government entities. PCEPI also comprises a broader range of goods and services. The Federal Reserve sometimes prefers to use PCEPI over CPI because the inflation rates based on PCEPI are less affected by short-time price changes.

CPI can be generally decomposed into three components: energy, food, and its core components, where the core CPI is equivalent to CPI less food and energy (Bauer et al., 2004 and references therein). It is well known that the energy and food components tend to have large and persistent price changes. The headline inflation rate, which is most widely used, and calculated as the change of CPI, is also more volatile than the core inflation rate, calculated as the change of CPI core. Similarly, PCEPI can be generally decomposed into its energy, food, and core components. Alternatively, PCEPI can be decomposed into durable goods and durable services.

In this paper, we propose to investigate the relationship between inflation rates and the market excess return. We intend to answer: 1) Do inflation rates have good predictive power for the market excess return? 2) If they do, which kind of inflation rates have the highest predictive power, aggregate, energy, food, or core? CPI or PCEPI? 3) Which methods may have good out-of-sample predictability? Single-factor, multi-factor linear models, or complex neural network/deep learning models? 4) What happens when including or excluding the latest COVID sample period?

We use the most updated monthly data from Amit Goyal's website (Goyal and Welch, 2008; Goyal et al., 2021) until December 2022 to obtain the market excess return (MERET) as the S&P 500 index return subtracted by the risk-free rate. The CPI and PCEPI data are obtained from the St. Louis Federal Reserve Bank's (FRED) database similar to Clarida, Gali, and Gertler (2000) and Guo et al. (2022). Our data sample starts in January 1959 when all the data related CPI and PCEPI and their components became available. As the inflation rates in the latest COVID-19 period are extremely volatile, we examine both the full sample period until December 2022 and the sample period pre-Covid ending in March 2020, the day the World Health Organization (WHO) announced the start of the COVID-19 pandemic.

There is an ongoing debate on whether the stock return is predictable (e.g. Ang & Bekaert, 2007 and references therein). The classical work of Campbell and Cochrane (1999) used a single-factor consumption surplus ratio to predict the market excess return. The influential work of Goyal and Welch (2008) and Goyal et al. (2021) found poor predictability overall with each of the 29 predictors studied in the literature. The most recent working paper Guo et al. (2022) advocate using a three-factor multiple linear regression on year-over-year inflation rates, along with stock variance and the log-price-earnings ratio (LPE), finding promising out-of-sample performance (See the detailed literature review in Section 3).

In this paper, we decompose CPI and PCEPI into various components, and investigate the prediction of market excess return using various inflation rates. We also apply various methods: single-factor linear models with inflation rates alone; three-factor linear models with inflation rates, stock variance, and log-price-earnings ratio; and neural networks/deep learning models. We compare with the popular mean-combination model of Rapach et al. (2010) as the benchmark. We also evaluate the prediction performance both in-sample and out-of-sample. The out-of-sample performance is evaluated as the out-of-sample R^2 (Campbell and Thompson, 2008) benchmarking with the historical mean.

We find that the relationship between various inflation rates and market excess return is always negative. This is consistent with earlier findings such as Marshall (1992), Bodie (1976), and Chen, Roll, and Ross (1986). If only inflation rates are considered as a single predictor for market excess return, the energy and food components of the CPI and PCEPI and the service component of PCEPI are significant at the 0.1 level while the rest are insignificant. However, in the three-factor linear regression models with various inflation rates, along with stock variance, and LPE, all the regression coefficients become significant at the 0.01 level.

Furthermore, we find that the inflation rates of the food component of the CPI have the highest in-sample adjusted R^2 among all the three-factor linear regression models with various inflation rates, stock variance, and LPE. If a single-factor linear regression model with inflation rates alone is used, the energy component of the CPI has the highest adjusted R^2 in-sample and the highest out-of-sample R^2, 0.88%, for month-ahead-prediction.

Finally, we find that a neural network model using year-over-year inflation rates based on the food component of CPI, stock variance, and LPE has the highest out-of-sample R^2 for the one-month-ahead prediction for market excess returns, 4.74% for pre-COVID sample until March 2020 and 5% for the full sample until December 2022. It consistently overperforms the benchmarking mean-combination model, which also has positive out-of-sample R^2 throughout the sample period. This confirms that there is predictive power over the simple historical mean. For the last ten years since 2010, the three-factor multiple linear model even outperforms the neural network models.

This paper has the following sections. Section 2 introduces the data we use and defines all the variables. Section 3 reviews the most relevant literature and summarizes our contributions. Section 4 presents various methods we use to predict the market excess returns and benchmark with the current state-of-the-art method. Section 5 presents our main results and findings. Section 6 concludes our work.

DATA

All the monthly data are constructed from the most recent updates until December 2022 from Amit Goyal's website (Goyal and Welch, 2008; Goyal et al., 2021), and the St. Louis Federal Reserve Bank's (FRED) database, spanning from January 1959 to December 2022.

Market Excess Return (MERET): The market excess return is defined as the S&P 500 index return subtracted by the risk-free rate from Goyal's updated website data. The continuously compounded returns on the S&P 500 index, with dividends, are calculated from the Center for Research in Security Press' (CRISP) end-of-month index values. The risk-free rate is the Treasure-bill rate, which is the three-month Treasury bill secondary market rate from the FRED database.

Inflation Rate (INF): Inflation is the price index such as Consumer Price Index (CPI) or Personal Consumption Expenditures Price Index (PCEPI), obtained from the FRED database. Inflation rate is the change in inflation. We mainly use the year-over-year log changes in the price index with one month report lag as in Guo et. al. (2022). For example, to calculate the year-over-year inflation rates at month t, INF_t, based on CPI, we use $INF_t = \ln\left(\frac{CPI_{t-1}}{CPI_{t-1-12}}\right)$, where CPI_{t-1} is lagged month t-1 CPI over CPI_{t-1-12}, with one-month lag twelve months prior. For robustness check, we also use the month-over-month log changes in the price index as in Goyal and Welch (2008).

INF_CPI: CPI Inflation Rate as the change in the CPI.

INF_PCEPI: PCEPI Inflation Rate as the change in the PCEPI.

CPI: The monthly Consumer Price Index for All Urban Consumers: All Items in U.S. City Average (CPIAUCNS), not seasonally adjusted. Abbreviation in parenthesis is the code used in the FRED database.

CPI_Energy: The monthly Consumer Price Index for All Urban Consumers: Energy in U.S. City Average (CPIENGNS), not seasonally adjusted.

CPI_Food: The monthly Consumer Price Index for All Urban Consumers: Food in U.S. City Average (CPIUFDNS), not seasonally adjusted.

CPI_Core: The monthly Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average (CPILFENS), not seasonally adjusted.

PCEPI: The monthly Personal Consumption Expenditures: Chain-type Price Index, seasonally adjusted.

PCEPI_EGS: The monthly Personal Consumption Expenditures: Energy Goods and Services (DNRGRG), seasonally adjusted.

PCEPI_Food: The monthly Personal Consumption Expenditures: Food (DFXARG), seasonally adjusted.

PCEPI_DG: The monthly Personal Consumption Expenditures: Durable Goods (PCEDG), seasonally adjusted.

PCEPI_Services: The monthly Personal Consumption Expenditures: Durable Services (PCES), seasonally adjusted.

PCEPI_Core: The monthly Personal Consumption Expenditures Excluding Food and Energy (PCEPILFENS), seasonally adjusted.

Stock Variance (SVAR): The stock variance is the scaled options-implied market variance calculated as in Guo et. al. (2022). It is the realized stock market variance based on daily S&P 500 market returns before January 1986, and the month-end options-implied stock variance after January 1986.

SVAR_GW: The stock variance according to Goyal and Welch (2008). It is the realized stock market variance computed as the sum of squared deviations of daily returns on the S&P 500 from CRISP.

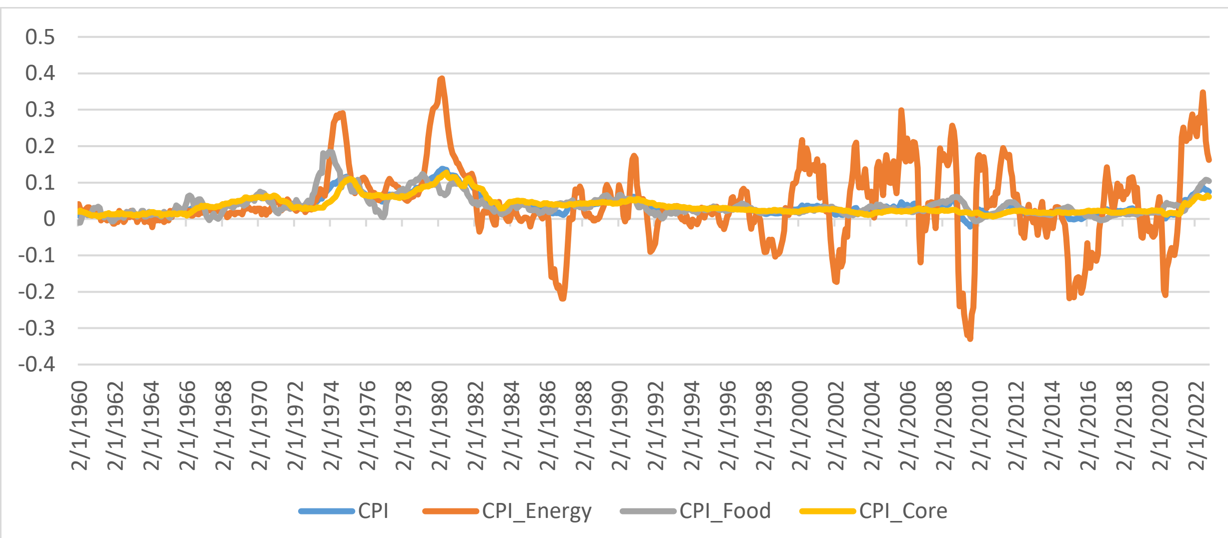
LPE: The logarithm of the price-earnings ratio as the scaled market price.

Table 1 gives the summary statistics for the full sample period from January 1959 to December 2022. The variables are the market excess return (MERET) and the year-over-year inflation rates of the following price indexes and their components: the whole consumer price index (CPI); the energy component of the CPI (CPI_Energy); the food component of the CPI (CPI_Food); the core components of the CPI (CPI_Core); the whole personal consumption expense price index (PCEPI); the energy component of the PCEPI (PCEPI_Energy); the food component of the PCEPI (PCEPI_Food); and the core components of the PCEPI (PCEPI_Core). PCEPI can be further decomposed into its durable goods component (PCEPI_DGS) and its services component (PCEPI_Services).

Table 1: Summary Statistics of Market Excess Return and Various Inflation Rates.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
MERET	-0.2179	-0.0180	0.0095	0.0066	0.0340	0.1648
CPI	-0.0212	0.0174	0.0295	0.0365	0.0459	0.1376
CPI_Energy	-0.3298	-0.0007	0.0270	0.0413	0.0871	0.3861
CPI_Food	-0.0102	0.0179	0.0283	0.0368	0.0458	0.1847
CPI_Core	0.0061	0.0195	0.0272	0.0361	0.0460	0.1274
PCEPI	-0.0148	0.0161	0.0250	0.0321	0.0417	0.1097
PCEPI_Energy	-0.3384	-0.0003	0.0253	0.0414	0.0857	0.3827
PCEPI_Food	-0.0191	0.0119	0.0232	0.0318	0.0424	0.1741
PCEPI_Core	0.0062	0.0162	0.0223	0.0316	0.0427	0.0973
PCEPI_DGS	-0.2103	0.0263	0.0582	0.0618	0.1010	0.5823
PCEPI_Services	-0.1791	0.0526	0.0671	0.0693	0.0922	0.2122

Figure 1: Year-over-Year Inflation Rates of CPI & Components.

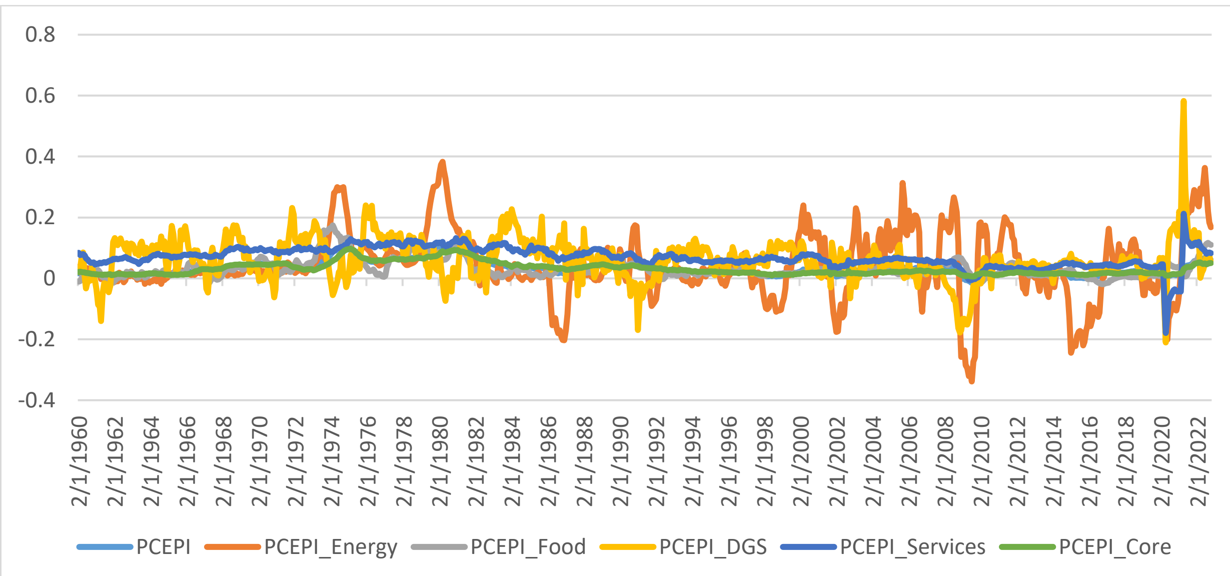


And Figure 1 shows the year-over-year inflation rates of CPI, and its components CPI_Energy, CPI_Food, and CPI_Core, or CPI less food and energy. We can see CPI_Energy appears most volatile with the peak at 38.61% and dip at -32.98%. Food is the second most volatile with maximum 18.47% and minimum -1.02%. On the other hand, the max of CPI is 13.76% and min of CPI is -2.12%; and the max of CPI_Core is 12.74% and min of CPI core is 0.61%. The CPI_Core excluding energy and food components appears the least volatile.

Figure 2 shows the year-over-year inflation rates of PCEPI, and its components PCEPI_Energy, PCEPI_Food, and PCEPI_Core, or PCEPI less food and energy. PCEPI can also be decomposed into PCEPI_DGS, durable goods and services, and PCEPI_Services. Similarly, PCEPI_Energy appears very volatile with the peak at 38.27% and dip at -33.84%. Food is the second most volatile with maximum 17.41% and minimum -1.91%. On the other hand, the max of PCEPI is 10.97% and min at -1.48%; and the maximum of PCEPI_core is 9.73% and minimum is 0.62%. For the decomposed PCEPI_DGS, and PCEPI_Servies, the maxima are 58.23% and 21.22% while the minima are -21.03% -17.91%. The PCEPI_Core excluding energy and food components appears least volatile among all the inflation rates based on PCEPI with maximum at 9.73% and minimum at 0.62%.

Figure 3 shows the correlation of market excess returns versus year-over-year inflation rates based on CPI, PCEPI, and their components. We can see that the associations between MERET and INF are negative but small in magnitude, less than -0.1. The correlation of inflation rates of CPI_Energy and PCEPI_Energy CPI and PCEPI, CPI_Food and PCEPI_Food, CPI_Core and PCEPI_Core are all high at 1.00, 0.98, 0.98, 0.97, respectively. On the other hand, the correlation between inflation rates of energy versus food and core of CPI and PCEPI are generally much lower, less than 0.46.

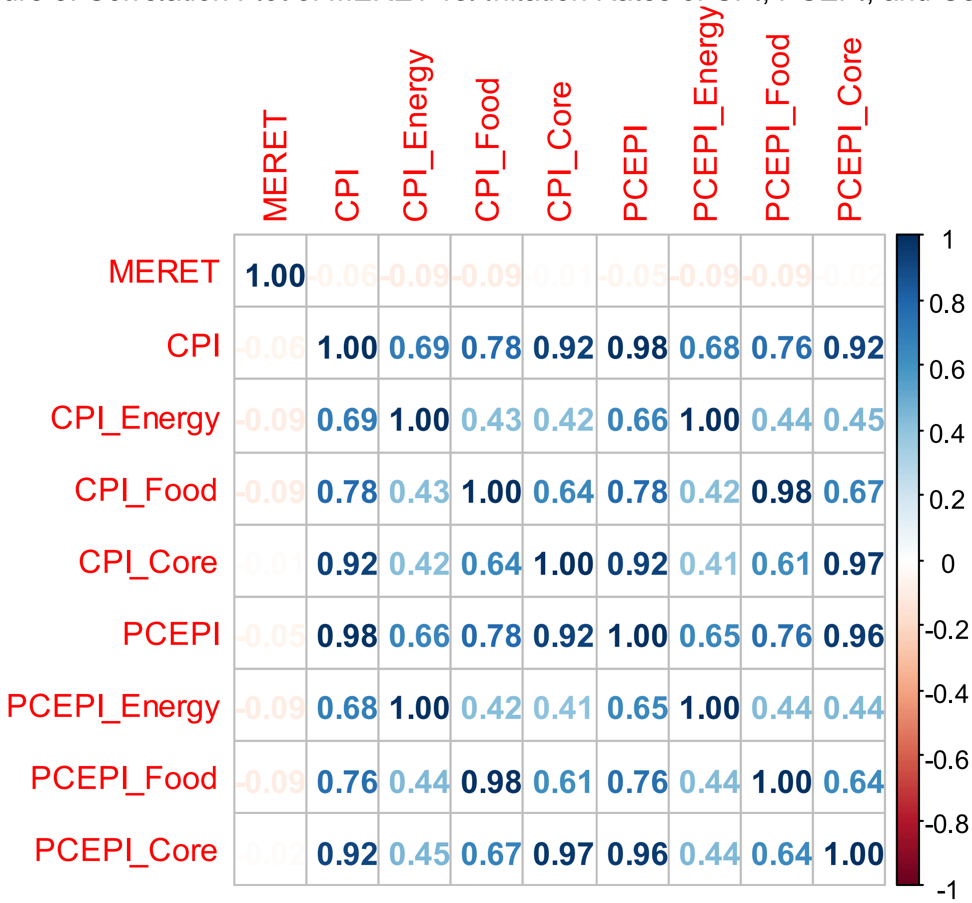
Figure 2: Year-Over-Year Inflation Rates of PCEPI & Components.



Full Sample vs. Pre-COVID Period

We will further present the analysis for the full sample and pre-COVID period in Section 5 Results. Since the WHO announced the COVID-19 pandemic on March 11, 2020, the inflation rates have become much more volatile, as exhibited in Figures 1 and 2 and Table 1. For example, in May 2021 during the COVID-19 pandemic, the year-over-year inflation rates in terms of durable goods and services, PCEPI_DGS, peaked as high as 58.23%; and in July 2021, CPI_Energy takes its second highest value at 34.80%. The associations of various measures of inflation rates with market excess returns before COVID-19 may be different from their relationships including the latest COVID-19 period. For this reason, we take a closer look into both the full sample that spans from January 1959 to December 2022 and pre-COVID period from January 1959 to March 2020.

Figure 3: Correlation Plot of MERET vs. Inflation Rates of CPI, PCEPI, and Components.



LITERATURE REVIEW

Extensive research has been conducted to search for the best market equity premium model. Table 1 provides a summary of the most recent research. Campbell & Thompson (2008) used a linear regression with weak constraints on the signs of coefficients, producing small, but meaningful, predictive power.

Goyal and Welch (2008) are an influential paper that investigated a comprehensive set of variables one-at-a-time to predict the market excess returns. They mostly focus on a 30-year sample period from 1975-2005 with data in monthly frequency. They find that most models they consider predict poorly both in-sample and out-of-sample.

Goyal, Welch, and Zafirov (2021) revisited the empirical performance of equity premium prediction through a comprehensive set of 29 predictors used in the previous literature. They also investigated machine learning methods. They concluded with the finding that predictive performance overall was still disappointingly poor.

Rapach et al. (2010) proposed the use of mean-combination models, investigating their predictive abilities by averaging the predictions from each single-factor model. They found that these models consistently provide significant predictive power relative to the historical mean; hence, we use this as our benchmarking model.

Atanasov et al. (2020) proposed a new consumption-based variable: cyclical consumption (CC). They found a negative relationship between future expected stock returns and aggregate consumption, while also finding some predictability.

Guo et al. (2022) conducted a comprehensive variable selection and find three predictors: year-over-year inflation rates, stock variance, and LPE are the most consistently selected important variables. They find that a three-factor multiple linear model performs even better than complex nonlinear machine learning models out-of-sample.

Table 2: Summary of Literature Review.

YEAR	REFERENCES	JOURNAL	MAIN METHOD	SAMPLE PERIOD	MAIN FINDING
2008	Campbell & Thompson	Review of Financial Studies	Linear regression with constraints on the signs of coefficients	Monthly, various til 2005	"by imposing restrictions….out-of-sample explanatory power is small, but … economically meaningful"
2008	Goyal and Welch	Review of Financial Studies	Single-factor, 17 variables	Monthly 1947:M1-2005:M4	"Predicted poorly both in-sample and out-of-sample"
2021	Goyal and Welch	SSRN	Single-factor, 29 variables	Various till 2020:M12	"Predictive performance remains disappointing"
2010	Rapach et al.	Review of Financial Studies	Mean-combination	Quarterly, 1947:Q1-2005:Q4	"Combining delivers statistically and economically significant out-of-sample gains relative to the historical average consistently."
2020	Atanasov et al.	Journal of Finance	Habit formation model using cyclical consumption based on PCE	Quarterly 1953:Q4-2017:Q4	Future expected stock returns are high (low) when aggregate consumption falls (rises). "The predictive power of cyclical consumption is not confined to bad times and subsumes the predictability of many popular forecasting variables."
2022	Guo et al.	SSRN	Three-factor multiple linear regression through variable selection	Quarterly, 1965:Q1 -2020:Q4	"The multifactor model is strikingly stable across subsamples with an out-of-sample R^2 of 9.6% over the 1965Q1 to 2020Q4 period, outperforming machine learning models and the combination forecast method."

We contribute to the literature by zooming in on inflation rates by dissecting the price indexes CPI and PCEPI into energy, food, and core components respectively. We investigate the relationship between various inflation rates and the market excess returns in a single-factor linear model, three-factor linear model along with stock variance, and LPE, and neural network/deep learning models. We also conduct a separate study on the full sample and pre-COVID sample. Interestingly, we find the inflation rates based on the food component of CPI along with stock variance and LPE using a neural network model can have as high as 5% out-of-sample R^2 for month-ahead prediction. If a single-factor model is used, the inflation rate of energy component of CPI alone gives the best out-of-sample performance.

METHODS

We explore the prediction of market excess return via single-factor linear model; three-factor linear model with inflation rates, stock variance, log price earnings ratio; and neural networks/deep learning models. We compare with the popular mean-combination model of (Rapach et al., 2010) as the benchmark. We also evaluate the prediction performance both in-sample and out-of-sample. The out-of-sample performance is evaluated as the out-of-sample R^2 (Campbell and Thompson, 2008) benchmarking with the historical mean.

Single-Factor Linear Model

We first examine the relationship between inflation rates and market excess return by fitting a single-factor linear regression model.

$$r_t = \alpha + \beta_1 INF_{t-1} + \epsilon_t. \tag{1}$$

Here r_t is the market excess return at month t, INF_{t-1} is the lagged predictor variable inflation rates at month t-1, and ϵ_t is the random error term. α is the constant or intercept term, and β_1 is the regression coefficient. We investigate the roles of inflation rates using CPI, PCEPI, and their components respectively. We focus on year-over-year inflation rates (Guo et al., 2022) as they have better predictive power than month-over-month inflation rates (Goyal and Welch, 2008). Our full sample spans from January 1959 to December 2022. Due to the calculation of year-over-year inflation with one-month reporting lag and one-more lag for one-month-ahead prediction, our effective number of monthly observations are 754 in model (1).

Three-Factor Linear Model

We then predict month-ahead market excess return using three predictors: inflation rates, stock variance, and log-price-earnings ratio. Again, inflation rates of using CPI, PCEPI, and their components are used along with SVAR and LPE. Our three-factor linear model to predict market excess return is:

$$r_t = \alpha + \beta_1 INF_{t-1} + \beta_2 SVAR_{t-1} + \beta_3 LPE_{t-1} + \epsilon_t. \qquad (2)$$

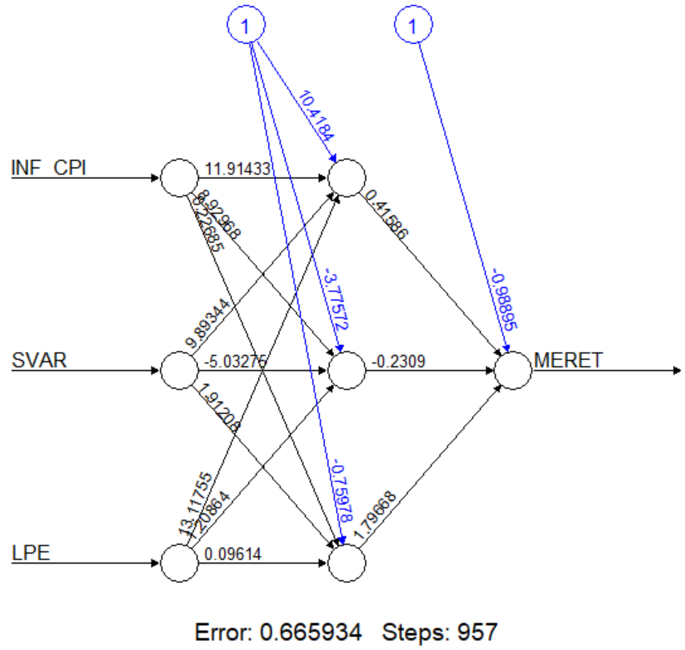
Similarly, r_t is the market excess return at month t, $INF_{t-1}, SVAR_{t-1}, LPE_{t-1}$ are the lagged predictor variables inflation rates, SVAR, and LPE at month t-1, and ϵ_t is the random error term. α is the constant or intercept term, and $\beta_1, \beta_2, \beta_{31}$ are the regression coefficients. We investigate the roles of inflation rates using CPI, PCEPI, and their components respectively along with SVAR and LPE. We focus on SVAR calculated as in Guo et al. (2022) where they use month-end option-implied stock variance after January, 1986 as they have better predictive power than realized variance as in Goyal and Welch (2008).

Artificial Neural Networks/Deep Learning

Neural network or deep learning are popular machine learning models that can incorporate complex nonlinearity and interactions (e.g. Hastie et al., 2009).

Let us illustrate the neural networks/deep learning methods we use through graphs. Figure 4 shows neural network plot of one hidden layer and three neurons (NN1(c(3))) for the full sample period February 1959 till December 2022.

Figure 4: Neural Networks NN1(c(3)) of Full Sample.



A neural network model has an input layer, hidden layer(s), and an output layer. In our case, the input layer consists of an input matrix of three predictors: $\boldsymbol{x}_{t-1} = [INF_{t-1}, SVAR_{t-1}, LPE_{t-1}]$. Let us first demonstrate a simple neural network with a single hidden layer and three neurons, and their corresponding weights $w_{i,j}^{(0)}, i = 1,2,3; \text{j} = 1, \dots, 3$ or weight vector $\boldsymbol{w}_1^{(0)}, \boldsymbol{w}_2^{(0)}, \boldsymbol{w}_3^{(0)}$ and the bias $b_{i,0}, i = 1,2,3,$ where i is the index for neuron, j is the index for predictor, superscript in parenthesis is the index for layer. For example, the first neuron in the hidden layer transforms the input layer into

$$z_{1,t-1} = f_1\,(\boldsymbol{w}_1^{(0)}\boldsymbol{x}_{t-1} + b_{1,0}) = f_1\left(w_{1,1}^{(0)}INF_{t-1} + w_{1,2}^{(0)}SVAR_{t-1} + w_{1,3}^{(0)}LPE_{t-1} + b_{1,0}^{(0)}\right).$$

Here f_1 is an activation function. We use a common activation function rectified linear unit (ReLU) which is defined as $f_1(u) = \max(\mathrm{u}, 0)$. Activation functions help model non-linearity and allow the network to model complex relationships. Besides ReLU, other common activation functions are sigmoid, tanh, and softmax.

Finally, the three neurons in the single hidden layer are linearly weighted into the output layer mathematically as follows:

$$r_t = h\left(\boldsymbol{w}_1^{(1)}z_{1,t-1} + \boldsymbol{w}_2^{(1)}z_{2,t-1} + \boldsymbol{w}_3^{(1)}z_{3,t-1} + b_{1,0}^{(1)}\right) + \epsilon_t,$$

Again, h is an activation function and ϵ_t is a random error term.

We can see from Figure 4, with this simple set-up of a single hidden layer with 3 neurons, the total number of parameters to estimate is 16=(3+1)*3 + 3+1, where 4=3+1 parameters of 3 weights and 1 bias for each of the 3 neurons in the hidden layer and 3 weights and 1 bias term to convert the hidden layer into the single output.

Deep learning is a synonym for neural networks with multiple hidden layers. It becomes a state-of-the-art machine learning tool, especially for image and speech recognition. Figure 5 shows a neural network of two hidden layers with 6 neurons in hidden layer 1 and 3 neurons in hidden layer 2 that we fit on our full data sample till December 2022.

We can see that the total number of parameters to estimate increases to 49=(3+1)*6 +(6+1)*3 3+1, where 4=3+1 parameters of 3 weights and 1 bias for each of the 6 neurons in hidden layer 1, 6 weights and 1 bias term in hidden layer 2, and 3 weights and 1 bias term to convert hidden layer 2 into the single output.

The tuning parameters of neural networks include the number of hidden layers, the number of neurons in each hidden layer, activation functions, learning rate, regularization parameters, batch size, and dropout rate etc. The number of hidden layers defines the depth of the neural network and influences its capacity to model complex patterns. The number of neurons also affects the network's ability to model complex relationships and interactions.

A deeper neural network and/or a larger number of neurons can incorporate more complex relationship, but it can significantly increase the number of parameters to estimate, and increase computational burdens. A machine learning principle is the tradeoff between model complexity and prediction accuracy, and tradeoff between bias and variance. A highly complex neural network may fit the in-sample data very well but possibly do not predict well, commonly known as the phenomenon of “overfitting”.

We use R package neuralnet() for our computation. We also implemented keras() and tensorflow(), and the results are comparable. We further note that as neural networks are highly nonlinear with many parameters to estimate, it is subject to local solution and can be sensitive to starting values. We use random starting points multiple times and use the neural network with the best objective value. We also implemented some packages with automatic tuning parameter selection. Our limited study shows that the computation time significantly increases for recursively out-of-sample prediction and the improvement of accuracy is limited. For this reason, we present results with neural networks with 1) one hidden layer, 3 neurons, denoted as c(3); 2) two hidden layers, 3 neurons for each, denoted as c(3,3); and 3) two hidden layers, 6 and 3 neurons respectively, denoted as c(6,3). We use ReLU as the activation function. And the rest of tuning parameters set as default options.

Mean-Combination Model

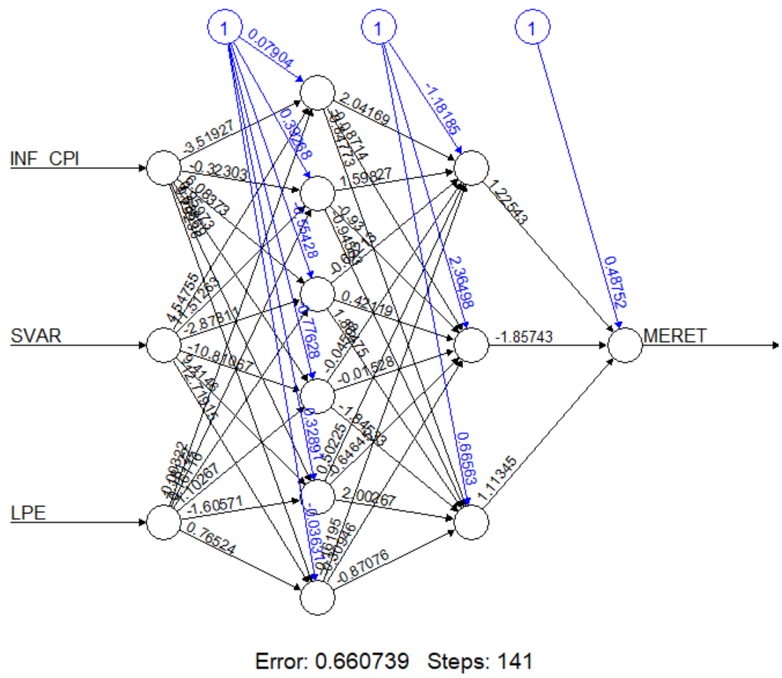
The mean-combination model of Rapach et al. (2010) is simply to run simple linear regression for each of the predictors and then average the predicted market excess returns. Guo et. al. (2022) show that the three-factor multiple linear regression outperforms the mean-combination model with 15 predictors in Rapach et al. (2010). We use the mean-combination model with three factors INFL, SVAR, and LPE as the benchmark model, which can be mathematically written as the following:

$$
\begin{aligned}
r_{INF,t} &= \alpha_{INF} + \beta_{INF} INF_{t-1} + \epsilon_t,\\
r_{SVAR,t} &= \alpha_{SVAR} + \beta_{SVAR} SVAR_{t-1} + \epsilon_t,\\
r_{LPE,t} &= \alpha_{LPE} + \beta_{LPE} LPE_{t-1} + \epsilon_t,
\end{aligned}
$$

$$r_{LPE,t} = \alpha_{LPE} + \beta_{LPE} LPE_{t-1} + \epsilon_t,$$
$$\hat{r}_t = (\hat{r}_{INF,t} + \hat{r}_{SVAR,t} + \hat{r}_{LPE,t})/3. \quad (3)$$

Here $\hat{r}_t$ is the one-month-ahead predicted market excess return. $\hat{r}_{INF,t} = \hat{\alpha}_{INF} + \hat{\beta}_{INF} INF_{t-1}$ is the predicted MERET based on a simple linear regression on month t-1 inflation rate and so on.

Figure 5: Neural Network c(6,3) of Full Sample.



RESULTS

In-Sample Estimates

We show in Table 3 the in-sample single-factor linear regression (SLR) results to predict monthly market excess returns (MERET) using year-over-year Inflation rates based on the Consumer Price Index (CPI), Personal Consumption Expenditures Price Index (PCEPI), and their components. The full sample spans January 1959 to December 2022. Due to the construction of year-over-year inflation rates with one-month lag, the effective lagged inflation rates start on February 1960 in the prediction model. Inflation rates (INF) based on CPI, CPI_Energy, CPI_Food, CPI_Core, PCEPI, PCEPI_EGS (Energy Goods and Services), PCEPI_Food, PCEPI_DG (Durable Goods), PCEPI_Services, and PCEPI_Core are used as the single-factor predictor respectively. We report the estimated coefficients, the corresponding t-values (in parenthesis), and the adjusted R square values (R_a^2). Significance levels at 0.01, 0.05, and 0.1 are denoted as ***, **, and *.

All the regression coefficients of various inflation rates are negative, which is consistent with the literature that higher inflation rates, the lower market excess returns are. The following variables were found to be significant at the 0.1 level: CPI_Energy, CPI_Food, PCEPI_EGS, PCEPI_Food, and PCEPI_Services. Both the CPI and PCEPI energy component were found to have the highest R_a^2 value of 0.0072, with the food components of the CPI and PCEPI having R_a^2 values just under at 0.0069 and 0.0071, respectively. On the other hand, the CPI and PCEPI with all components and the core components exhibit small, or even negative, R_a^2 values, with insignificant regression coefficients, indicating poor predictability even while in-sample.

Table 3: Single-factor linear model results of market excess return on various inflation Rates.

#	CPI	CPI_Energy	CPI_Food	CPI_Core	PCEPI	PCEPI_EGS	PCEPI_Food	PCEPI_DG	PCEPI_Services	PCEPI_ Core	R^2_a
1	-0.0942 (-1.617)										0.0021
2		-0.0385 (-2.536)*									0.0072
3			-0.129 (-2.503)*								0.0069
4				-0.0201 (-0.308)							-0.0012
5					-0.0978 (-1.446)						0.0014
6						-0.0377 (-2.540)*					0.0072
7							-0.126 (-2.522)*				0.0071
8								-0.0066 (-0.288)			-0.0012
9									-0.0966 (-2.005)*		0.0040
10										-0.0417 (-0.54)	-9E-04

In Table 4, we show the in-sample three-factor linear regression (MLR) results to predict monthly market excess returns (MERET) using year-over-year inflation rates based on the Consumer Price Index (CPI), Personal Consumption Expenditures Price Index (PCEPI), and their components along with stock variance (SVAR) and log-price-earnings ratio (LPE). The full sample spans January 1959 to December 2022. Due to the construction of year-over-year inflation rates with one-month lag, the effective lagged inflation rates start on February 1960 in the prediction model. Inflation rates (INF) based on CPI, CPI_Energy, CPI_Food, PCEPI, PCEPI_EGS (Energy Goods and Services), PCEPI_Food, PCEPI_DG (Durable Goods), and PCEPI_Services are used as the inflation-rate predictor respectively.

Table 4: In-sample Three-Factor Regression on Year-over-Year Inflation Rates of CPI, PCEPI, and their Components along with SVAR and LPE of the Full Sample.

#	CPI	CPI_Energy	CPI_Food	PCEPI	PCEPI_EGS	PCEPI_Food	SVAR	LPE	R^2_a
1	-0.385 (-4.576)***						1.139 (4.763)***	-0.0242 (-4.446)***	0.0392
2		-0.0616 (-3.664)***					0.8614 (3.778)***	-0.0128 (-3.076)***	0.0297
3			-0.305 (-4.950)***				1.1955 (4.971)***	-0.018 (-4.060)***	0.0436
4				-0.487 (-4.644)***			1.152 (4.808)***	-0.027 (-4.613)***	0.0399
5					-0.05976 (-3.652)***		0.8807 (3.774)***	-0.0127 (-3.057)***	0.0296
6						-0.289 (-4.926)***	1.217 (5.023)***	-0.0168 (-3.907)***	0.0433

We find that the three-factor linear model with inflation rates of CPI_Food along with SVAR and LPE has the highest R_a^2 at 0.0436, much higher than the R_a^2 of 0.0072 from the best single-factor linear model with energy components in CPI and PCEPI. In the three-factor linear model, all the regression coefficients of the various inflation rates stay negative but are now significant at the 0.01 level. Similarly, all the regression coefficients of the SVAR and LPE variables are also significant at the 0.01 level. As expected, all the SVAR regression coefficients are positive and all the LPE regression coefficients are negative, which is consistent with the literature.

Out-of-Sample Predictions

We adopt the common measure of out-of-sample R^2, denoted as R_{OOS}^2 (Campbell and Thompson 2008) to assess the out-of-sample performance of various models.

$$R_{OOS}^2 = 1 - \frac{\sum_{t=t_0+1}^{T}(r_t - \hat{r}_t)^2}{\sum_{t=t_0+1}^{T}(r_t - \bar{r}_t)^2} \qquad (4)$$

Here $\hat{r}_t$ is the predicted market excess return at time t based on one-month-prior information available at t-1, $\bar{r}_t$ is the simple rolling mean prediction based on historical average over all the past market excess return observed one month ago. We start the first out-of-sample prediction at time month 301 using the first 25-year or t_0 =300-month observations for in-sample estimation and then use the expanding window. That is, we next use 301-months observed data to predict one-month-ahead market excess return at month 302 and so on. Our sample period starts on the first available lagged predictors of various year-over-year inflation rates of CPI, PCEPI, and their decomposed components, SVAR, and LPE in February 1960. The last market excess return to predict is at the time point December 2022.

A model performs well in-sample may not necessarily predict well out-of-sample. It is well known that a complex model can often fit the training data well but not the testing data, a phenomenon known as “overfitting”. The out-of-sample R_{OOS}^2 in equation (4) measures the decrease in the mean squared prediction error of one-month-ahead MERET using the single-factor model based on various inflation rates alone; multi-factor (3-factor) model; neural networks/deep learning models; and the mean-combination model, all with INF, SVAR and LPE, compared to that of simply using historical average. If R^2_{OOS} is positive, then the predictions performs better than the simple average. The higher positive value of R^2_{OOS} is, the better predictive power is for the market excess return.

Table 5: Out of Sample R^2_{OOS} of various Model Fits of Full Sample.

Method(INF+SVAR+LPE)	CPI	CPI_Energy	CPI_Food	PCEPI	PCEPI_EGS	PCEPI_Food
MLR	0.0305	0.0185	0.0241	0.0290	0.0172	0.0277
NN1(c(3))	0.0456	0.0198	0.0302	0.0453	0.0103	0.0407
NN2(c(3,3))	0.0330	0.0169	0.0046	0.0037	0.0170	0.0062
NN3(c(6,3))	0.0290	0.0096	0.0500	0.0337	0.0179	0.0357
Combination	0.0089	0.0117	0.0104	0.0084	0.0117	0.0111
SLR (INF)	0.0042	0.0088	0.0065	0.0030	0.0086	0.0081

Table 5 presents the one-month ahead expanding window out of sample R^2_{OOS} of various three-factor model fits with inflation rates, stock variance, and log price-earnings ratio and a single-factor simple linear regression with inflation rates of the full sample from January 1959 to December 2022. All the R^2_{OOS} are shown to be positive, confirming predictive power over historical mean. The three-factor multiple linear regression model (MLR) appears to consistently have more predictive power than the benchmarking mean combination model. Out of all the combinations of inflation rates in the MLR model, the one including CPI is shown to have the highest R^2_{OOS} of 0.0305.

In the neural network models presented, the simplest one hidden-layer three-neuron neural network model (NN1(c(3))) performed well overall, while the more complex two hidden-layers neural network model with six neurons in the first layer and three in the second layer (NN3(c(6,3))) has the highest R^2_{OOS} in the table of 0.0500 with the CPI food inflation rates (CPI_Food).

The single linear regression model with only inflation rates shows positive R^2_{OOS} values, with the highest being 0.0088 with the CPI energy inflation rate (CPI_Energy), which is comparable with the mean combination model with the inflation rate of the whole CPI and SVAR and LPE.

Table 6: Out of Sample R^2_{OOS} of various Model Fits of Pre-COVID Sample.

Method(INF+SVAR+LPE)	CPI	CPI_Energy	CPI_Food	PCEPI	PCEPI_EGS	PCEPI_Food
MLR	0.0294	0.0161	0.0253	0.0286	0.0143	0.0253
NN1(c(3))	0.0430	0.0141	0.0299	0.0404	0.0108	0.0299
NN2(c(3,3))	0.0250	0.0094	0.0062	0.0037	0.0102	0.0062
NN3(c(6,3))	0.0312	0.0032	0.0474	0.0343	0.0121	0.0062
Combination	0.0081	0.0102	0.0098	0.0078	0.0102	0.0098
SLR (INF)	0.0018	0.0046	0.0048	0.0010	0.0043	0.0048

Table 6 presents the one-month ahead expanding window out of sample R^2_{OOS} of various three-factor model fits with inflation rates, stock variance, and log price-earnings ratio and a single-factor simple linear regression model with only inflation rates of the pre-COVID sample from until March of 2020. Again, all the R^2_{OOS} values are positive, showing predictive power over the historical mean. The MLR still consistently has more predictive power than the benchmarking mean combination model. Out of all the inflation rates in the MLR model, the one including CPI is still the highest, with an R^2_{OOS} value of 0.0294.

In the neural network models presented, the same pattern as in Table 5 is seen, with the NN1(c(3)) model performing better overall but the more complex NN3(c(6,3)) model again attaining the highest R^2_{OOS} value in the table of 0.0474 with the CPI_Food inflation rate.

The single linear regression model with only inflation rate still has only positive R^2_{OOS} values, with the highest being 0.0048. The SLR model is thus incomparable in predictive power with the rest of the model methods.

Figure 6: Out-of-Sample R^2_{OOS} of sample period until March 2020, pre-COVID, with predictors CPI inflation rates, SVAR, and LPE.

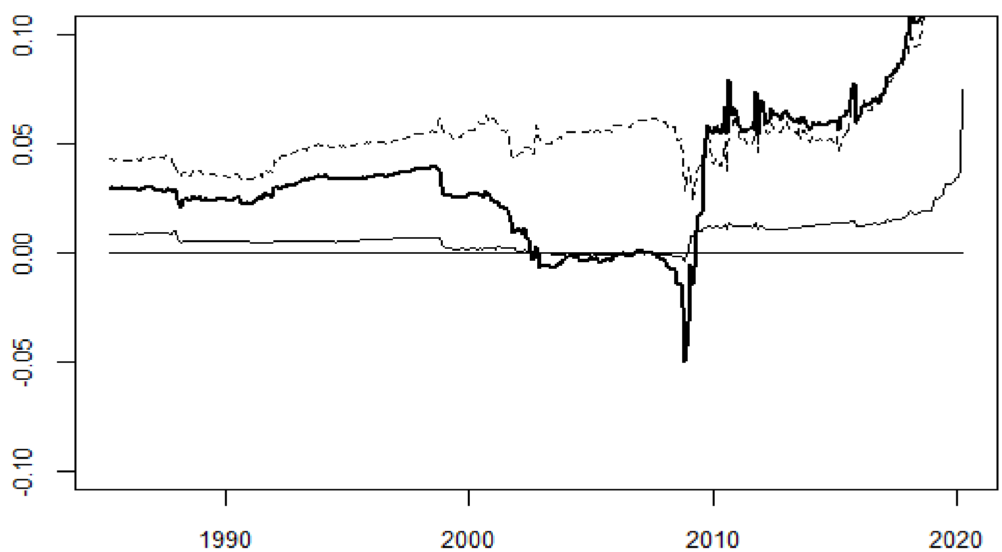


Figure 6 plots the out-of-sample R^2_{OOS} of three-factor multiple linear regression (solid line), neural network model with one hidden layer and three neurons, c(3), (dashed line), and the mean combination model (gray line). The three predictors are INF_CPI, SVAR, and LPE. We can see that the R^2_{OOS} of the mean combination model are positive in general, confirming predictive power in Rapach et al. (2010). The c(3) neural network model also gives positive R^2_{OOS} and performs consistently better than the mean-combination model over the whole out-of-sample period. The three-factor multiple linear regression model performs the best for the latest 10-year period post January 2010. Its R^2_{OOS} is positive except for a negative dip around October 2009.

Figure 7: Out-of-Sample R_{OOS}^2 of sample period until March 2020, pre-COVID, with predictors CPI_Food inflation rates, SVAR, and LPE.

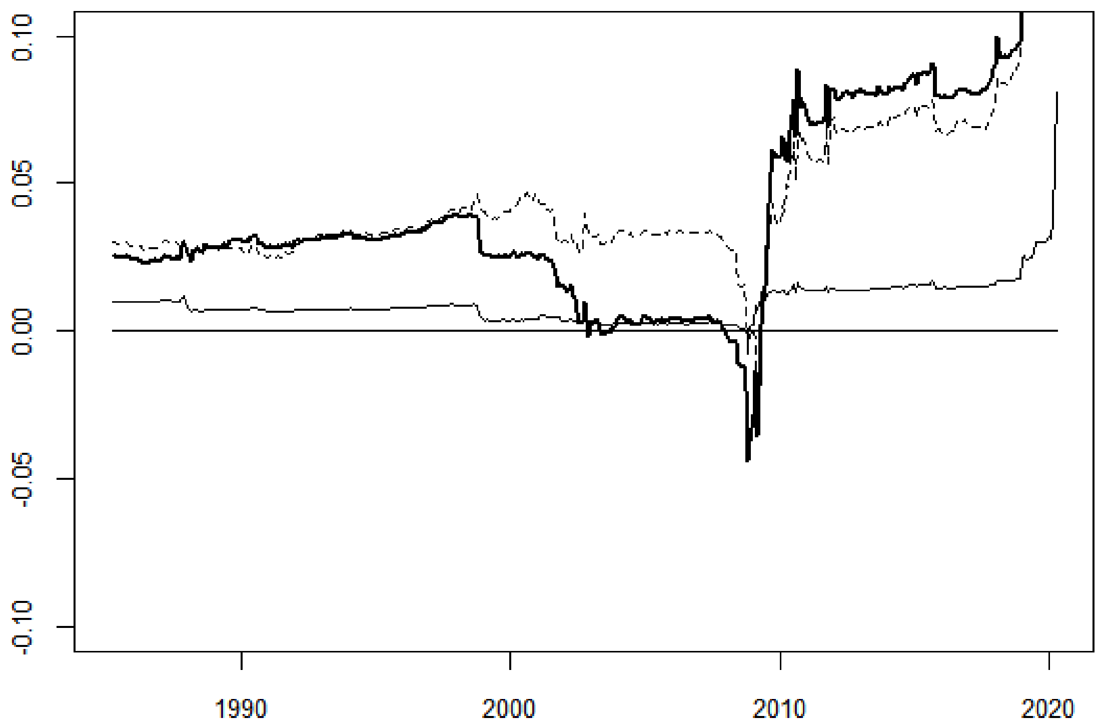


Figure 7 plots the out-of-sample R_{OOS}^2 of three-factor multiple linear regression (solid line), the c(3) neural network model with one hidden layer and three neurons (dashed line), and the mean combination model (gray line). The predictors are the CPI_Food inflation rate, SVAR, and LPE. We see that the mean combination model is positive, confirming its predictive power as in Rapach et al. (2010). Both the c(3) neural network model and the three-factor multiple linear regression model have positive R_{OOS}^2 , except for the dip in October 2009, and consistently out-perform the mean combination model over the whole out-of-sample period. We find once again that the three-factor multiple linear regression performs the best in the latest ten-year period after January 2010.

Figure 8: Out-of-Sample R^2_{OOS} of sample period until March 2020, pre-COVID, with predictors PCEPI inflation rates, SVAR, and LPE.

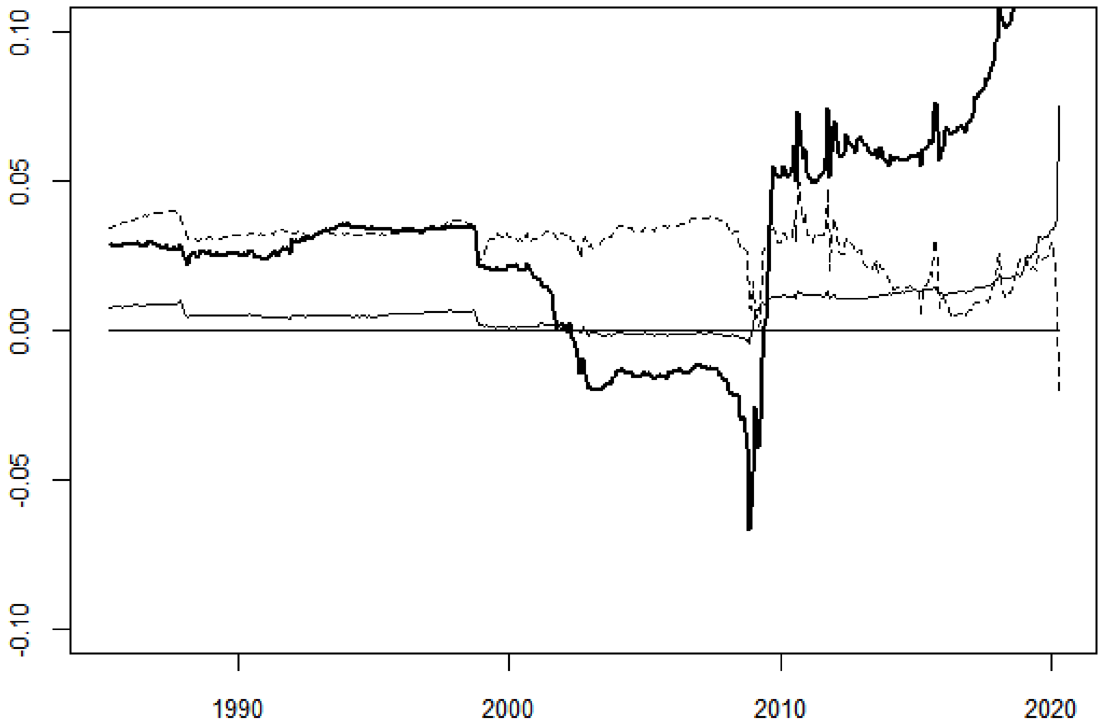
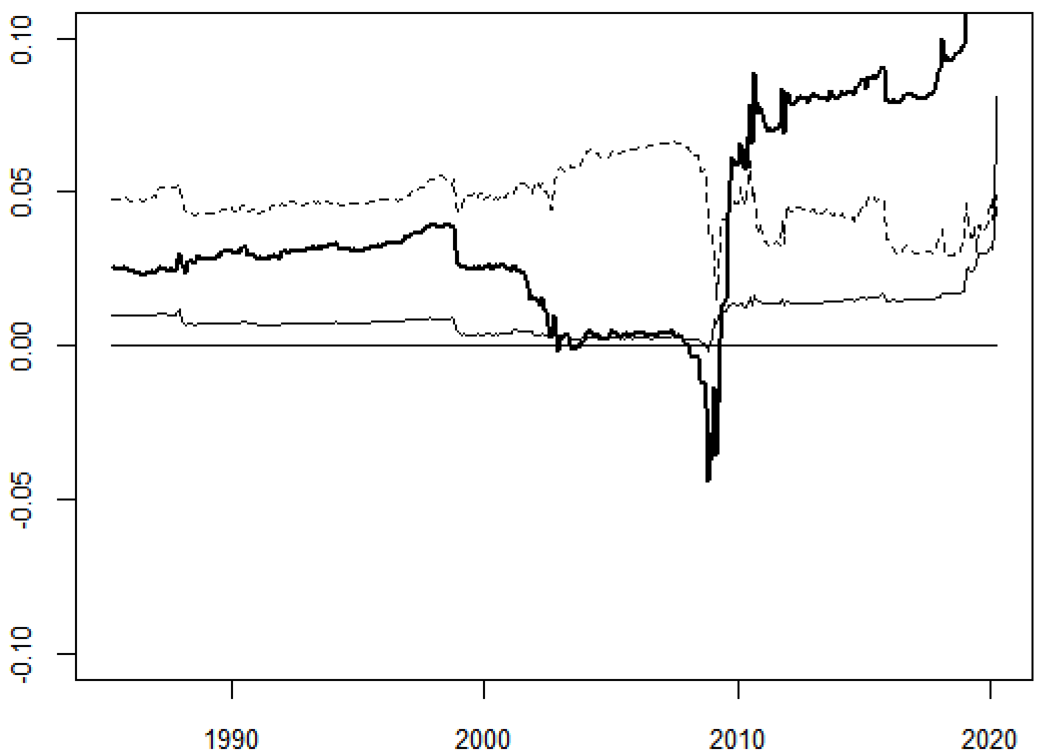


Figure 8 plots the out-of-sample R^2_{OOS} of three-factor multiple linear regression (solid line), the c(3) neural network model with one hidden layer and three neurons (dashed line), and the mean combination model (gray line). The predictors are the PCEPI inflation rate, SVAR, and LPE. We see that the mean combination model is positive, confirming its predictive power as in Rapach et al. (2010). Both the c(3) neural network model and the three-factor multiple linear regression model have positive R^2_{OOS} , except for the dip in October 2009, and consistently out-perform the mean combination model over the mid-period. We find once again that the three-factor multiple linear regression performs the best in the latest ten-year period after January 2010. In comparison, using the CPI inflation rates is better in terms of out-of-sample prediction, which is also evidenced in Table 6.

Figure 9: Out-of-Sample R^2_{OOS} of sample period until March 2020, pre-COVID, with predictors CPI_Food inflation rates, SVAR, and LPE and NN3(c(6,3)) neural network model.



From Table 6, using the NN3(c(6,3)) neural network model with CPI_Food inflation rates has the highest R^2_{OOS} of 0.0474. As evidenced in Figure 9, all three curves are generally positive, showing their predictive power. The neural network model appears to perform well in early and mid-sample, while the three-factor multiple linear regression performs best in the latest ten-year period.

DISCUSSION AND CONCLUSIONS

We zoom in and dissect consumer price index (CPI) and personal consumption expenditures price index (PCEPI) into the energy, food, and core components to measure various inflation rates and investigate their roles in predicting market excess returns (MERET). We find negative relationship between all inflation rates and market excess returns. The single-factor regression coefficients of inflation rates alone are only significant at 0.1 level for energy, food, and services components and the coefficients of CPI and PCEPI alone are even insignificant. However, the three-factor regression coefficients of all inflation rates along with stock variance and LPE are significant at 0.01 level. We find strong out-of-sample predictive power based on the inflation rates of food components of CPI along with stock variance and LPE. Among several neural network models we use, the inflation rates of food components of CPI along with stock variance and LPE can have the highest out of sample R_{OOS}^2 as 5% for one-month-ahead prediction of the S&P 500 excess returns.

For future work, it is worth studying how tuning parameters and starting values influence complex neural network/deep learning models. Gu et al. (2020) study at firm level stock excess return and then aggregate to the market excess return both cross-sectionally and in time series. This may also be worthwhile exploring in the future. For the ongoing mixed findings of stock return predictability, we hope this empirical study can help with a better understanding of how various inflation rates of the whole price indexes and their components are related to and their predictive power to the market excess return.

REFERENCES

Ang, A. & Bekaert, G. (2007). Stock return predictability: Is it there? *The Review of Financial Studies*, 20, 651-707.

Bauer, A., Haltom, N., & Peterman, W. B. (2004). Decomposing inflation. *Economic Review*, Federal Reserve Bank of Atlanta, 89, 39-51.

Bodie, Z., (1976). Common stocks as a hedge against inflation, *The Journal of Finance,* 31, 459-470.

Campbell, J. Y. & Cochrane, J. H. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy*, 107, 205-251.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4), 1509-1531.

Chen, N., Roll, R., & Ross, S. A. (1986), Economic forces and the stock market, *Journal of Business*, 383-403.

Clarida, R., Gali, J., & Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory, *The Quarterly Journal of Economics,* 115, 147-180.

Goyal, A. & Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455-1508.

Goyal, A., Welch, I., & Zafirov, A. (2021). A comprehensive look at the empirical performance of equity premium prediction II. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3929119

Gu, S., Kelly B., & Xiu, D. (2020), Empirical asset pricing via machine learning, *Review of Financial Studies*, 33, 2223-2273.

Guo, H., Sanni, S., & Yu, Y. (2022). Searching for the best conditional equity premium model, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4168294

Hastie, T., Tibshirani, R., & Friedman, J.H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. New York: Springer.

Marshall, D. A., (1992). Inflation and asset returns in a monetary economy, *The Journal of Finance,* 47, 1315-1342.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2), 821-862.